Identifying optimal number of driving cycles to represent diverse driving conditions

Asad Yarahmadi, Catherine Morency, and Martin Trépanier

Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, Montréal, QC, Canada · Interuniversity Research Centre on Enterprise Networks, Logistics, and Transportation (CIRRELT), Montréal, QC, Canada · Chair in Transportation Transformation, Montréal, QC, Canada · Department of Mathematical and Industrial Engineering, Polytechnique Montréal, Montréal, QC, Canada

Published in: *International Journal of Sustainable Transportation*, 2024, Vol. 18, No. 8, pp. 704–726. DOI: 10.1080/15568318.2024.2397647. Received 20 November 2023; Revised 20 August 2024; Accepted 21 August 2024.

Abstract

Driving cycle is one of the main inputs of vehicle emission modeling. However, the variability of driving cycles due to fluctuations in weather conditions is one of the primary sources of uncertainty in vehicle emission estimation. This study aims to identify and determine an optimal number of driving cycles that can correctly represent driving patterns in diverse weather conditions. First, a multivariate multiple regression model is developed to determine the most important weather factors affecting the driving patterns. Then, similar weather conditions are identified according to these factors using unsupervised machine learning. Next, two driving cycles are constructed for diverse weather types, one for weekdays and one for weekends. Afterward, descriptive analysis and a similarity matrix are employed to determine how similar the generated driving cycles are in different weather types. Finally, 15 driving cycles are identified to represent driving patterns in diverse driving conditions.

Keywords: Driving behaviors variability; driving cycle; vehicle emissions; weather conditions

1. Introduction

Vehicle emissions are a major contributor to air pollution and climate change, making it a critical issue affecting the health and well-being of humans and the environment. As the world becomes increasingly dependence on transportation, it is imperative to find ways to resolve the existing uncertainty of estimation of the harmful emissions produced by cars. The exactitude of tailpipe emissions largely depends on the accuracy of the driving cycle.

The driving cycle represents the average driving behaviors in a region and can be characterized by a set of driving speed, acceleration, and braking patterns. It has been widely employed in designating fuel consumption (Achour & Olabi, 2016; Ma et al., 2019), assessing vehicle performance (Degraeuwe & Weiss, 2017; Ladubec et al., 2021), optimizing EMS (Energy Management System) (Chen et al., 2019; Wu et al., 2023), and estimating driving style (Rios-Torres et al., 2019). Driving cycles most importantly associate drivers' behavior with exhaust emissions in vehicle emissions models like MOVES (MOtor Vehicle Emissions Simulator) (Bodisco & Zare, 2019; EPA, 2014; Muhammad, 2022).

Numerous studies have examined the influences of the main weather-based parameters, such as precipitation, temperature, wind, and visibility, on drivers' behaviors (Jägerbrand & Sjobergh, 2016; Kilpeläinen & Summala, 2007). Findings indicate that motorists often adapt their speeds in response to changing weather conditions, which in turn affects fuel consumption (Abdi Kordani et al., 2018; Andrey et al., 2013; Chen et al., 2019; Chowdhury, 2015; Christian & Jensen, 2014; Elsagheer Mohamed et al., 2022; Faria et al., 2017; Zahid et al., 2023; Zhang et al., 2018). However, the analysis of weather's impact on the driving cycle parameters has been infrequently explored by researchers. With the increase in global temperatures and climate change, choosing the sets of driving cycles that best reflect the weather conditions and the resulting vehicle emissions is essential.

Generally speaking, there are two types of driving cycle including: Standard Driving cycles (SDC) and Local Driving Cycle (LDC) (Gebisa et al., 2021). A thorough analysis of existing knowledge reveals that the existing driving cycle production approaches are beset with two major flaws. First, previous investigations tend to inadequately incorporate adverse weather conditions, such as heavy rainfall and severe winds, into the process of collecting driving data and constructing driving cycles. For example, SDCs predominantly consider climatic parameters that affect the engine's thermodynamic efficiency, often disregarding adverse weather conditions that may affect the drivers' behavior. Similarly, Worldwide Harmonized Light-duty Vehicle Test Procedure (WLTP) methodology prioritizes temperature and humidity (United Nations, 2014), the Canadian standard driving cycle, 5-cycle test, incorporates a restricted range of temperatures (–7 °C, 35 °C, and 20–30 °C), the USA FTP-75 test scrutinizes a narrow ambient temperature spectrum 20–30 °C (Eckert et al., 2018). Second, prior studies have commonly utilized one or few driving cycles to represent driving behaviors across different meteorological conditions, neglecting the variations in driving behaviors under diverse environmental scenarios. For instance, Ma et al. (2019) synthesized two driving cycles, the off-peak and peak hours, for Beijing. Chauhan et al. (2020) developed one driving cycle to mimic driving behaviors for Vadodara City. Kaymaz et al. (2019) constructed one driving cycle to represent driving patterns in Istanbul. This oversimplification potentially engenders uncertainties in the estimation of vehicle emissions.

To consider the influences of variability of weather conditions on driving behavior, there are two principal challenges. First, weather conditions exhibit significant variability over the course of a day and year, and these changes can impact driving patterns. For instance, in the city of Montréal, Canada, the temperature can fluctuate from −30 °C to 35 °C throughout the year and over 20 °C in a single day, while humidity ranges from 20% to 100% annually. Conversely, the reactions of drivers to these weather conditions depend on a variety of factors, including demographic characteristics (such as age and gender), vehicle attributes (such as type and condition), and driving skills (such as experience and training) (Yeo et al., 2021). Due to the complex interplay of these factors, identifying the number of representative driving cycles that accurately capture the driving behaviors volatility influenced by weather patterns is a challenging task. Second, the arduous process of generating driving data to encompass all driving scenarios and a diverse range of drivers is time consuming and cost prohibitive. As a result, the produced LDCs very often rely on inadequate data, obtained within a limited timeframe and involving a restricted number of drivers (Al-Samari, 2017), as evidenced by studies conducted by (Ameknassi et al., 2016; Cui et al., 2022; Kazemi Miyangaskari et al., 2023; Kazemi Miyangaskary et al., 2023; Zhai et al., 2018).

The principal objective of this study is to devise an approach for identifying the optimal number of representative driving cycles that capture driving patterns under different weather conditions. Establishing a diverse array of driving cycles under various weather conditions can enhance the accuracy and reliability of vehicle emissions estimates, which are crucial for developing robust environmental policies and advanced vehicle technologies. The significance of this study extends beyond traditional applications, as it provides critical insights into how shifting weather patterns, underpinned by climate change, affect driving behaviors. Such knowledge is essential for anticipating changes in vehicle performance and emissions in a future marked by increasingly unpredictable weather conditions. To achieve the desired objective, this study collected comprehensive driving and weather datasets to cover various drivers' behaviors and different weather conditions.

As weather conditions affect driving cycle variables, the number of driving cycles in a region can be a function of number of weather types. Therefore, it is crucial to disentangle the effects of weather factors from those of road features on driving cycle parameters (Abele & Møller, 2011; Ragione & Giovanni, 2016). This is achieved by collecting driving data on a specific road type selected based on the findings of our precedent study (Yarahmadi, 2023; Yarahmadi et al., 2023).

This study examines the following question:

- How many driving cycles are required to represent driving behaviors in different weather conditions?
- What weather parameter(s) have the most significant impact on driving cycle parameters?

It presents three notable contributions aimed to advance the understanding and optimization of vehicular performance:

• First, a systematic approach is proposed to establish a statistically significant relationship between weather conditions and the attributes of driving cycles.

• Second, a methodology is proposed to generate and identify the optimal quantity of driving cycles for diverse driving scenarios, thereby facilitating the development of efficient and effective transportation systems.

• Third, the impacts of weather parameters on driving cycle characteristics are isolated from the influence of road features.

The remainder of this article is structured as follows. First, previous studies that explored the relationship between driving conditions and driving cycles are discussed, and their arguments are exposed. Then, the methodology section describes the technical steps to find optimal sets of driving cycles. After, the article presents the optimal sets of driving cycles, and their various parameters are elaborated. Finally, the results are discussed, and some perspectives are proposed.

2. Related work

The current section reviews the research problem and identify the knowledge gap. To do that, first, the conventional approach of incorporating the impact of weather conditions in the development of driving cycles is reviewed. Second, the current practices for selecting the most representative driving cycle are discussed in detail.

The SDC is a process approved and applied by authorities to evaluate and certify a model of vehicle that complies with all performance standards (relative to emissions) before authorizing it to be introduced to the market (European Commission Joint Research Center, 2013). SDCs are commonly utilized to evaluate and compare vehicle emissions across various regions, whether at a tertiary or global level. In order to ensure their applicability and relevance for different regions, SDCs must be designed with general conditions in mind, encompassing a wide range of driving scenarios, traffic patterns, and environmental factors. This broad approach allows for a more inclusive and representative assessment of vehicle emissions, ultimately contributing to developing effective policies and standards for emission control worldwide. SDCs predominantly concentrate on the parameter of ambient temperature and do not inherently emulate other meteorological conditions like precipitation, wind velocity, humidity, or visibility. The principal justification for this approach is rooted in the necessity of establishing a regulated, reproducible setting for the comparative analysis of various vehicles. The incorporation of additional variables such as precipitation or wind speed into these assessments would introduce further fluctuations, complicating the task of contrasting vehicles under uniform, standardized conditions. Table 1 illustrates that the established SDCs often fail crucial weather factors that can impact vehicle emissions.

Table 1. SDCs and weather parameters.

Name of SDC	Country	Weather parameters
New European driving cycle	European Union	Temperature: 20–30 °C
WLTC	European Union United Kingdom Turkey, China, India	Temperature: (23 °C ± 5 °C) Temperature: –7 °C Relative humidity: 40%–60%
Highway Fuel Economy Test (HWFET)	United States	Temperature: 20–30 °C
Japanese JC08 Cycle	Japan	Temperature: 20 °C ± 5 °C
KDC	Korean	Temperature: 20–30 °C
India Drive Cycle (IDC)	India	Temperature: 25 °C average
CB	Brazil	Temperature: 20–30 °C
ADR 81/02	Australia	Temperature: 20 °C ± 5 °C

Multiple studies have underscored a discrepancy between emissions estimates derived from real-world data and those obtained through Standard Driving Cycles (SDCs). For example, the International Council on Clean Transportation (ICCT) reported a growing divergence in this disparity in Europe, from 8% in 2008 to 39% in 2017 (Greene et al., 2017). To address the limitations, Local Driving Cycles (LDCs) are developed for specific regions, incorporating real-world driving data to synthesize these cycles (Seers et al., 2015). As a result, LDCs provide more accurate estimations of vehicular emissions compared to Standard Driving Cycles (SDCs) (Anida et al., 2017; Esteves-Booth et al., 2001; Fotouhi & Montazeri-Gh, 2012).

However, acquiring a comprehensive dataset encompassing all variables influencing vehicular conduct for developing a representative local driving cycle necessitates a considerable investment of time and financial resources. Consequently, the researchers very often concentrated their efforts on capturing region-specific traffic attributes to ascertain representative driving cycle parameters utilizing limited data sets. At the same time, exogenous factors such as meteorological conditions are often neglected.

Liu et al. (2020) collected data for 15 consecutive days to establish a representative driving cycle for buses in China, which may not be sufficient to capture the variability in driving conditions. Galgamuwa et al. (2016) identified a representative driving cycle for an expressway in Sri Lanka by collecting 56 h of driving data over three days, which may not be comprehensive enough to account for different driving conditions. Pouresmaeili et al. (2018) gathered driving data for two weeks under clear weather and dry road conditions in Mashhad, Iran, to develop a representative driving cycle, but the limited scope of data collection may not consider the full spectrum of driving variability.

In addition, Simulation-based methodologies, while useful in controlled settings, may not accurately capture the complex interplay of factors present in real-world scenarios, such as the impact of weather conditions on driving dynamics. For example, the study by (Amirjamshidi & Roorda, 2013; Keivanpour, 2022), which utilized simulated data to construct driving cycles for light and heavy trucks, failed to incorporate the influence of meteorological factors on driving behavior, thus limiting the ecological validity of their findings. As a result, the LDCs constructed for the same region have varied drivers' behavior representation.

Furthermore, the second gap is related to the authors' technique for selecting representative driving patterns. Indeed, unlike the first group, some studies gathered the required data in different seasons but eventually picked only one driving cycle as a representative.

Ashtari et al. (2014) believe that insufficient data and ambient conditions are two factors that generate uncertainty in vehicle emissions. Accordingly, they gathered 44 million driving data points for one year to develop a driving cycle in Winnipeg, Canada. Finally, they invented a new snipped selection algorithm to find the best candidate driving cycle. Ma et al. (2019) developed 17,000 driving cycles but selected two as the most representative driving cycle for peak and off-peak hours. Günther et al. (2017), over one full year, collected around 27,365 h of driving operation data in Hamburg, Germany, to produce driving cycles for public buses. Based on this vast data, the authors developed a driving cycle.

In conclusion, two gaps in the literature can be discussed. First, preexisting driving cycle generation methods do not consider the impact of weather parameters on driving cycle characteristics. They mostly account for a limited number of weather conditions when collecting or constructing driving cycles. This issue is particularly evident in Local Driving Cycles (LDC), where researchers often use small driving datasets to construct these cycles. Second, most studies select only one or a few driving cycles to represent driving behaviors under all driving conditions.

3. Research methodology

This segment presents the data utilized in the present investigation, elucidating the methodological involved. Additionally, Figure 1 visually delineates the stages undertaken during the course of this study. This study encompasses a comprehensive six-step approach, explained in detail in Subsections 3.2–3.5. As articulated, this research aims to determine the optimal number of driving cycles pertinent to diverse driving conditions. To that end, crucial data about weather and driving are gathered and prepared in the first and second steps. Subsequently, a sensitivity analysis identifies the critical meteorological factors significantly influencing driving patterns. Weather types are subsequently generated in the next step based on these factors. In the fifth stage, a sensitivity analysis is carried out to explore how the proposed model effectively captures the complex relationship between weather conditions and driving patterns. Finally, two driving cycles are generated for each weather type, corresponding to weekdays and weekends.

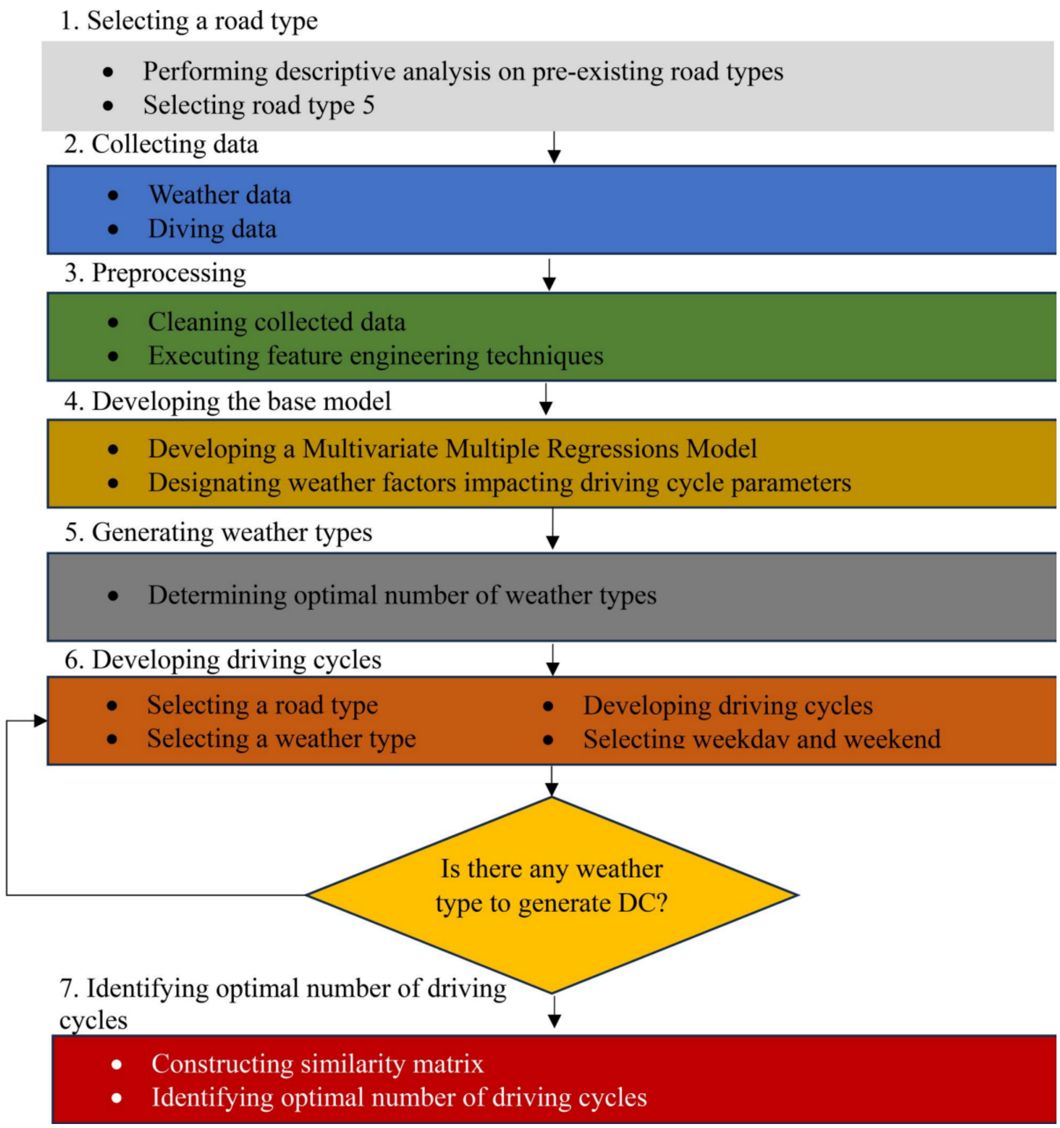


Figure 1. Research methodology of the present study.

3.1. Study area and data description

Prior to performing the present study, authors developed a road typology, detailed in Yarahmadi et al. (2023), involved classifying Montreal's road network into five types based on factors affecting driving speed. Subsequently, as illustrated in Figure 1, road type 5 is selected. The primary reason to focus on one road type is to isolate the impact of road features from the weather conditions impacts. This approach enables us to shed light on how changing driving conditions influences driving cycle parameters.

The main reason to select this road type 5 is because, among all road types, it has the highest number of segments, totaling 2608, with a combined length of almost 35 km. It features a diverse array of road characteristics, which are almost homogeneously distributed across the segments. Table 2 provides a detailed overview of the defining characteristics of the road segments within this selected category.

Table 2. Characteristics of the selected road type.

Road feature	Values	Proportion of values in the database
Road functions	Motorway	25.3
	Primary	26
	Secondary	43
	Tertiary	3.7
	Ramp	2
Land uses	City Center	4.1
	Industry	25.1
	Mixed Activities	27.7
	Residential area	39.1
	Park	4
Speed limit (km/h)	50	55
	60	10
	70	35
Number of Lanes	1	1
	2	24.5
	3	69
	4	5.5
Gradient (%)	–1	3.8
	0	41.2
	1	25
	2	18
	3	12
Density of signalized intersections	0	23.2
	1	21
	2	27.7
	3	14.1
	4	8
	5	6
Accessibility	Min: 0.1, Average: 0.42, Max: 1.5	–
Radius of curvatures (m)	Min: 102, Average: 455, Max: 1,587	–

3.1.1. Driving data

To conduct this study, a fleet of 3323 connected taxicabs produced a vast dataset consisting of approximately 3,227,000 trips undertaken between April 2019 and December 2020 under genuine driving conditions, including both peak and off-peak periods when the cabs were in service (only trips with passengers onboard are used). The characteristics of the driving data are presented in Table 3. Additionally, the driving data was not selectively gathered based on ancillary attributes such as the vehicle type, driver's age, or gender. Thus, this issue lets authors suppose the data likely originated from driving data belonging to various age groups, genders, and vehicle passenger types.

Table 3. Driving data features.

Variable	Timestamp	Speed	Longitude	Latitude
Type	Date	Number	Number	Number

3.1.2. Weather data

Montreal's climate is continental, with frigid winters and warm summers (Wood, 2015). Hourly basis weather data was collected from Environment Canada for 2019 and 2020. The main weather parameters are temperature, precipitation, wind speed, and humidity. It is worth mentioning that in this dataset with time-resolved data, snow and rain are reported collectively, and this study treats them as a single variable. Figure 2 indicates the variation of weather parameters over 2019 and 2020.

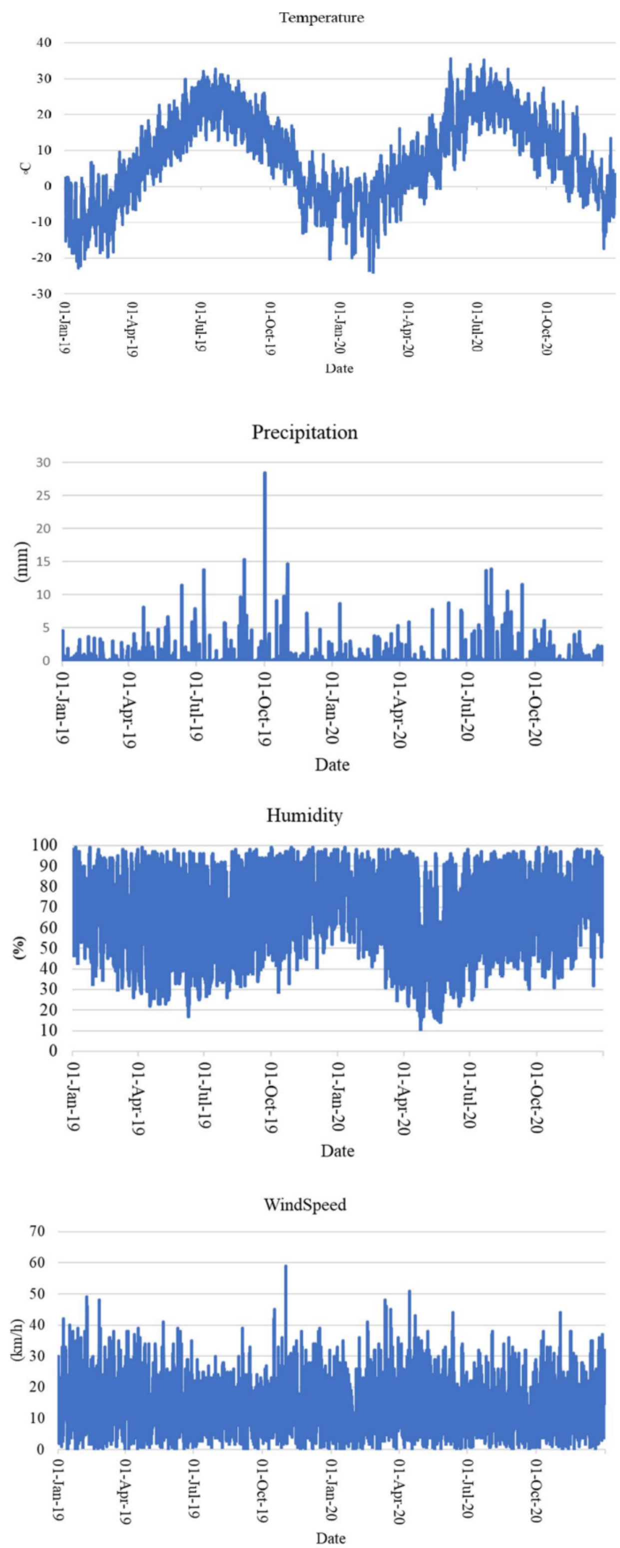


Figure 2. Variation of temperature, humidity, precipitation, wind speed over the years (2019–2020).

3.2. Data preprocessing

The second step of this study is data preprocessing. Outlier analysis and data cleaning techniques elucidate unnecessary data in both datasets. Afterward, two steps are taken to prepare the data: generating required driving data attributes (data enrichment) and feature engineering. Each phase is explained in the following sections.

3.2.1. Data enrichment

Data enrichment is a step to generate new items from existing data to make it ready for analysis. To create a driving cycle, driving data should be generated per second. It is because driving modes are generated based on a one-second interval. Investigating the raw data showed that data are mainly available in 5-s intervals or more. To address this issue, a resampling technique is used to increase the time interval granularity from 5 s to 1 s. However, while this adjustment increases the frequency of time stamps, it results in null values for other driving data parameters, such as driving speed, latitude, and longitude, at the newly created time stamps. Literature shows that the Kalman filter is a powerful tool to impute null values in time series data (Ribeiro, 2021). To validate the results obtained from the Kalman Filter, the imputed data were examined using QGIS. This process involved a comparative analysis where both the pre-imputation dataset and the post-imputation dataset, resampled at a one-second interval, were loaded into QGIS. By overlaying these datasets in the QGIS environment, I was able to visually assess the accuracy of the imputations. Specifically, I compared the imputed positions against known data points to evaluate the consistency and reliability of the imputed values. This approach allowed for a detailed examination of how well the Kalman Filter preserved the overall patterns and characteristics of the original data while filling in the gaps. In the final step, the distance between each point is calculated using the Geopy library in Python and the Haversine formula. The Haversine formula is a very accurate mathematical technique to calculate the distance between two points on the surface of a sphere.

The most critical weather factors are made available by Canada Environment, and there is no additional feature that the authors can calculate. Although visibility is another vital weather parameter that affects driving patterns, it is not available at the hourly level. Unavailability of some data is one of the limitations faced by this study.

3.2.2. Feature engineering

Feature engineering combines techniques to make input data compatible with machine learning algorithms and reduce computational complexity (Chen et al., 2014).

Regarding the weather data, the first feature engineering step is correlation analysis. As its name implies, the correlation between segment variables is explored in this step, and highly correlated variables, referred to as multicollinearity, are deleted. A feature is strongly correlated if its correlation coefficient is more than 0.7 with another part (Ratner, 2009). To perform the correlation analysis, the Dython library in Python is utilized. This library is based on the correlation ratio (Lewandowski et al., 2007). The output illustrates that none of the driving and weather data were correlated.

So, the second step of feature engineering was to evaluate the distribution of every segment feature. To do that, the skewness of features was measured. Skewness shows how symmetric data distribution is. Skewness can help identify potential issues in the data, such as outliers, non-normality, or extreme values. In addition, some clustering algorithms assume that the data follow a specific distribution, often a multivariate normal distribution. Assessing skewness can help to verify whether the data meet these assumptions and to select appropriate clustering techniques. If the skewness values of a column exceed 0.5, it is skewed. The SciPy library in Python programming is used to calculate the skewness. The result shows that the precipitation and wind speed were highly and slightly skewed respectively. To treat and analyze skewed data there are three methods comprising: implementing transformations, utilizing nonparametric methodologies, or synergizing robust statistical methods with analytical procedures. In the first step, transformation techniques are used to cope with skewness. The square root transformation yielded superior results for wind speed data, whereas the quantile transformation performed better for precipitation data. Despite this, the distribution of precipitation data remained skewed.

Another area for improvement that could affect the accuracy of machine learning technique output is the uniformity consistency of units of continuous unscaled data. For instance, the unit of precipitation is centimeters, the wind speed is m/s, and so on. To resolve this matter, all weather parameter values were standardized. To achieve this, each value is standardized by subtracting the mean value of that feature and then dividing by the standard deviation. As a result, the standardized values typically have a distribution with a mean of 0 and a standard deviation of 1.

Dimensionality reduction, the final feature engineering technique applied in this study, serves a critical purpose beyond simply identifying and omitting variables with minimal variance. It addresses the curse of dimensionality, where high-dimensional spaces can lead to sparse data distributions and make clustering analyses less effective due to increased computational complexity and difficulty in identifying meaningful patterns. Principal Component Analysis (PCA), a widely recognized dimensionality reduction technique, is particularly valuable in this context. PCA transforms the original variables into a new set of uncorrelated features, known as principal components, which are linear combinations of the original variables weighted by their contribution to the total variance. By constructing a covariance matrix and determining its eigenvectors and eigenvalues, PCA quantifies the variance each feature contributes to the dataset. The eigenvectors represent the directions of maximum variance, while the corresponding eigenvalues indicate the magnitude of variance along these directions. The selection of principal components is based on the cumulative explained variance ratio derived from these eigenvalues, ensuring that the reduced dataset retains the most significant variance within the data. This process not only streamlines the feature space, enhancing computational efficiency, but also improves the interpretability of clustering by focusing on the most informative aspects of the data. By mitigating noise and reducing overfitting risks, PCA allows for a more robust and meaningful clustering outcome that can reveal hidden structures within the data, which might otherwise be obscured in a higher-dimensional context. It is confirmed that the performance of PCA is reduced if data is not normally distributed. Conversely, Robust PCA (RPCA), one of the extensions of PCA, shows better performance in skewness cases (Candès et al., 2011). In the present study, PCA is used for parametric clustering techniques, while RPCA is used for non-parametric clustering techniques.

3.3. Multivariate multiple regression model

It is necessary to benefit from regression techniques to provide insight into the relationship between weather parameters and the driving cycle. In addition, its result can be interpreted to understand how the generated model can describe the relationship between independent and dependent variables. There are various regression models, but the main reasons for selecting multivariate multiple regression model (MMRM) are as follows:

Simultaneous analysis: Given the presence of 13 numerical dependent variables in the data, MMRM permits a simultaneous examination of the relationships between the multiple predictor and outcome variables. This allows us to holistically understand the interactions within the datasets.

Revealing complex relationships: MMRM is capable of uncovering intricate relationships and interactions among variables, which may be overlooked if separate univariate analyses are conducted. This results in a more nuanced understanding of the data.

Combined effects analysis: MMRM facilitates the investigation of combined or collective effects of predictor variables on the dependent variables. This provides insights not just into individual influences, but also into how these predictors might work together to influence the outcomes.

The weakness of the MMRM is that it is sensitive to outliers and collinearity. As described in Subsection 3.2.2, the features engineering techniques were employed to cope with these issues.

3.4. Cluster analysis

Clustering, a form of unsupervised learning, constitutes a robust analytical technique aimed at categorizing data into distinct groups based on shared attributes. By maximizing intra-cluster similarities and concurrently accentuating inter-cluster dissimilarities, this approach effectively delineates homogeneous subsets within heterogeneous datasets, providing valuable insights into the underlying structure and relationships among observations.

Clustering techniques can be tailored to different data types, such as parametric, nonparametric, and mixed datasets. Nonparametric approaches, such as Affinity Propagation make no distribution assumptions and can handle diverse data types, including non-Gaussian or skewed distributions. Although k-means is categorized as a nonparametric technique, its performance can be less than optimal for skewed data due to its assumptions about cluster shapes, sensitivity to outliers, and reliance on minimizing variance. For heterogeneous datasets comprised of both skewed and non-skewed attributes, two approaches are recommended for effective clustering: (1) employing robust preprocessing techniques to prepare the data, followed by clustering, and (2) utilizing nonparametric clustering methods. To do that, the three following approaches are taken:

- Preparing data by robust PCA and then performing K-means.
- Employing Affinity Propagation (AP)
- Agglomerative Hierarchical Clustering (AHC)

In this study, weather types serve as the foundational basis for generating driving cycles corresponding to each weather category. The accuracy of these generated weather types significantly impacts the precision of the derived driving cycles. To ensure a comprehensive understanding of the relationship between weather conditions and driving cycles, this research employs various techniques to generate diverse weather types, acknowledging that each method has its unique advantages and disadvantages. Utilizing multiple approaches allows for the generation of diverse weather types, which can be compared and contrasted to identify the most accurate and representative categories. Subsequent cluster validation helps ascertain the most reliable results, ensuring the credibility and relevance of the generated driving cycles.

Agglomerative Hierarchical Clustering (AHC) and Affinity Propagation (AP) clustering methods that do not require pre-specifying the number of clusters. k-means, conversely, is a fast, user-friendly algorithm that performs well with continuous variables. However, its primary limitation is the need to determine the number of clusters a priori, which can often result in incorrect assumptions by the user. To address this issue, various techniques can be employed to determine the optimal cluster count. The following section provides a concise overview of existing methods for this purpose, offering an understanding of the approaches available to determine cluster numbers.

3.4.1. Determining the optimal number of clusters

Establishing the ideal number of clusters is a critical aspect of clustering analysis, as it significantly influences the outcome's quality and the algorithm's ability to discern patterns. In addition, selecting the appropriate number of clusters contributes to mitigating overfitting and underfitting, increasing the comprehensibility of results, and optimizing computational efficiency. Striking this balance is essential for revealing concealed structures and patterns within the dataset, ultimately ensuring a reliable and robust model performance. Xu et al. (2016) documented various techniques to determine the number of clusters.

The well-acknowledged techniques to determine the optimal number of clusters are as follows:

- **Elbow method:** This method involves plotting the within-cluster sum of squares (WCSS) against the number of clusters and selecting the number of clusters at the "elbow" of the curve where adding more clusters no longer significantly reduces the WCSS.
- **Gap statistics:** This method compares the within-cluster dispersion for different numbers of clusters to a reference distribution and selects the number of clusters that maximizes the gap between them.

• A **dendrogram** is a tree-like diagram that illustrates hierarchical clustering by depicting the relationships among data points. It enables an intuitive understanding of the data's organization and helps determine the optimal number of clusters by examining linkage distances.

In the current research, the optimal quantity of distinct weather classifications is dependent on the ideal number of clusters; hence, all previously mentioned methodologies are employed to ascertain this number of clusters. Additionally, to determine the optimal "K" parameters, the authors took into account various criteria such as the size of the dataset, the nature of the data, the objectives of the analysis, and preexisting knowledge in the field.

3.5. Driving cycle construction

Driving cycles are constructed using variation of the driving speed over time. There are four main steps to generate a driving cycle: microtrip generation, clustering of the generated microtrips, constructing candidate driving cycles, and selecting the most representative driving cycle. These steps are further detailed below.

3.5.1. Microtrip generation

A microtrip is a small part of a trip, and the sum of all microtrips builds a trip. In the microtrip generation step, driving pattern parameters are calculated for each microtrip and are used to compare all microtrips.

There are various approaches to generating a microtrip. The most famous ones are as follows:

• **Trip-based microtrip:** microtrip is bounded by two stops.

• **Fixed time interval microtrip:** a trip is divided into time intervals.

• **Fixed distance microtrip:** a trip is divided based on a fixed distance.

• **Intersection microtrip:** a microtrip is bounded by two intersections.

Nouri and Morency (2017) generated driving cycles using the above-mentioned methods and then employed a comparative and quantitative evaluation to ascertain the most effective microtrip generation method. By constructing driving cycles via various techniques, each cycle was then assessed against target performance metrics. The difference between the actual and target values of these metrics was calculated for each cycle. Methods were ranked by the magnitude of their deviation from the target, with rankings aggregated across all metrics to determine an overall score. This comprehensive analysis revealed that the fixed distance method of 250 meters yielded the most accurate replication of real-world driving conditions, as it consistently ranked highest, indicating the smallest deviations from the target metrics. Thus, for the purposes of this study, the driving cycle was generated according to this method.

3.5.2. Microtrip clustering

After generating microtrips, the next step is to find similar driving patterns. To do that, a dissimilarity matrix is used to compare microtrips by calculating the chi-square distance between their speed-acceleration distribution matrices. Each matrix quantifies the frequency of various speed and acceleration combinations during a microtrip. The chi-square distance is computed by assessing the difference between the observed frequencies in the empirical data and the expected frequencies, if the speed and acceleration were independent. The sum of these differences for all combinations gives the chi-square distance. This distance forms the elements of the dissimilarity matrix, where a lower value indicates higher similarity between the driving behaviors of two microtrips. This matrix helps in ranking the microtrip generation methods based on their accuracy in capturing real-world driving dynamics. Then, the matrix is clustered using the clustering technique to identify and group similar driving patterns into a cluster.

3.5.3. Constructing candidate driving cycles

The driving cycle generation in this study involves a structured process using microtrips from a database of trips. Initially, each driving cycle begins with one of the starting microtrips, chosen to represent different starting conditions in the database—these do not reflect a cold start as they are not from an idle state.

Following the selection of the initial microtrip, subsequent microtrips are chosen based on a one-step Markov model. This model utilizes a transition matrix to calculate the likelihood of transitioning from one microtrip to another, based on observed sequences in the dataset. The selection process involves identifying a microtrip that follows the initial one with the highest probability, ensuring that the starting speed of the new microtrip is within 2 km/h of the ending speed of the previous one (Nouri, 2015).

This sequence continues, adding microtrips to build a complete driving cycle, with each addition evaluated against specific criteria to ensure continuity and representativeness of real-world driving conditions. The approach combines statistical probability with practical constraints on speed transitions to create realistic and coherent driving cycles. For each weather condition candidate driving cycles are constructed. The average length of driving cycles was 25 min.

3.5.4. Selecting the most representative driving cycle

All candidate driving cycles are ranked based on assessment criteria to select the most representative driving cycle. The assessment criteria used in this study are presented in Table 4. In addition, authors used these parameters to characterize the generated driving cycles and to investigate the association between weather factors and driving cycles.

The assessment criteria can be categorized into three groups: velocity indicators, acceleration and deceleration indicators, and driving mode indicators. Driving mode refers to the manner in which a vehicle is being driven at a given moment, as determined by its speed, acceleration, and deceleration patterns. Commonly recognized driving modes include a percentage of idling (when the engine is running, but the vehicle is stationary), acceleration (when the vehicle is speeding up), deceleration (when the vehicle is slowing down), cruise (when the vehicle is moving at a steady speed), and creeping (when the vehicle is moving at a very slow speed). Monitoring and analyzing driving modes can provide valuable insights into a range of factors related to vehicle performance, fuel efficiency, and environmental impact. To generate the driving cycle in this study, the application developed by Roy and Morency (2020) is used.

Table 4. Driving cycle assessment criteria.

Assessment criteria	Denoted	Unit
Difference in Speed Acceleration Frequency Distribution (SAFD)	NA	(%)
The average speed of the entire driving cycle	V	(Km/h)
Average running speed	Vr	(Km/h)
Maximum speed	Vm	(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	NA
Root means square acceleration	Acc2	(m/s^2)
Road power	NA	(KW)
Time proportion of idle mode	Idle_p	(%)
Time proportion of acceleration mode	Acc_p	(%)
Time proportion of cruising mode ($-0.1\ m/s^2$ < acceleration < $0.1\ m/s^2$; average speed > 5 m/s)	Cru_p	(%)
Time proportion of deceleration mode (acceleration ≤ $-0.1\ m/s^2$)	Dcc_p	(%)
Time proportion of creeping mode ($-0.1\ m/s^2$ < acceleration < $0.1\ m/s^2$; average speed < 5 m/s)	Cre_p	(%)

4. Results and discussion

In this section, the results of the developed regression models are assessed and discussed, and clustering approaches are implemented to recognize various weather types.

4.1. Sensitivity analysis

In pursuit of recognizing the influence of weather conditions on driving cycle parameters, attributes specified in Figure 2 and Table 3 are employed for a sensitivity analysis. As producing driving cycles for all hourly data is not possible, the initial step necessitates use of sampling techniques to extract a subset from the weather data.

The rationale for applying the method to a subset of weather data stems from the computational and practical challenges associated with generating driving cycles for each hour of the two-year period. This process is not merely data-intensive but also computationally demanding due to the complexity and granularity of the driving cycle generation process. Given these constraints, a representative sample was necessary to facilitate a feasible analysis while still capturing the variability of weather conditions over the selected period.

To ensure that selected samples offer a fair representation of the dataset throughout the year, the stratified sampling approach is adopted. This method facilitates breakdown of the population into homogeneous subsets, termed strata, and consequently allows for random selection of samples from each stratum. Thus, weather data is segregated into four distinct strata, each embodying a particular season. Consequently, for the longer seasons of winter and summer, we selected more samples. For the shorter seasons of spring and fall, fewer samples were chosen. In total, we selected 472 samples. In other words, we selected 472 h driving data in different weather conditions. Following these steps, driving cycles are synthesized for each hour and totally 472 driving cycles were constructed. Then, developed driving cycles and weather sample data are integrated according to their date and time. Eventually, MMRM model was created, and driving cycle parameters and weather factors are considered as dependent and independent variables respectively and weather factors as deterministic factors were designated. The R^2 of the generated model is 0.28.

4.2. Generating weather types

Clustering methods consider the influential factors outlined in Subsection 4.1 to categorize similar weather conditions. This study takes three distinct approaches to cluster weather variables, thereby producing diverse weather typologies.

4.2.1. Determining the optimal number of clusters

Three prevalent techniques—Elbow, Gap Statistics, and Dendrogram—were employed to ascertain the optimal number of clusters for the weather dataset. The Elbow and Gap Statistics methods were applied to the k-means algorithm, while the Dendrogram approach was utilized for AHC. As illustrated in Figure 3, all three techniques consistently identified 8 clusters as the ideal quantity for this dataset. In the Elbow method (a), the blue line represents the cluster numbers, and the dashed green line demonstrates the training time required for the clustering model per cluster number. In the Gap Statistics method (b), the optimal number of clusters is highlighted by a red circle. Lastly, in the Dendrogram method (c), the red line indicates the cut line utilized to determine the optimal cluster count.

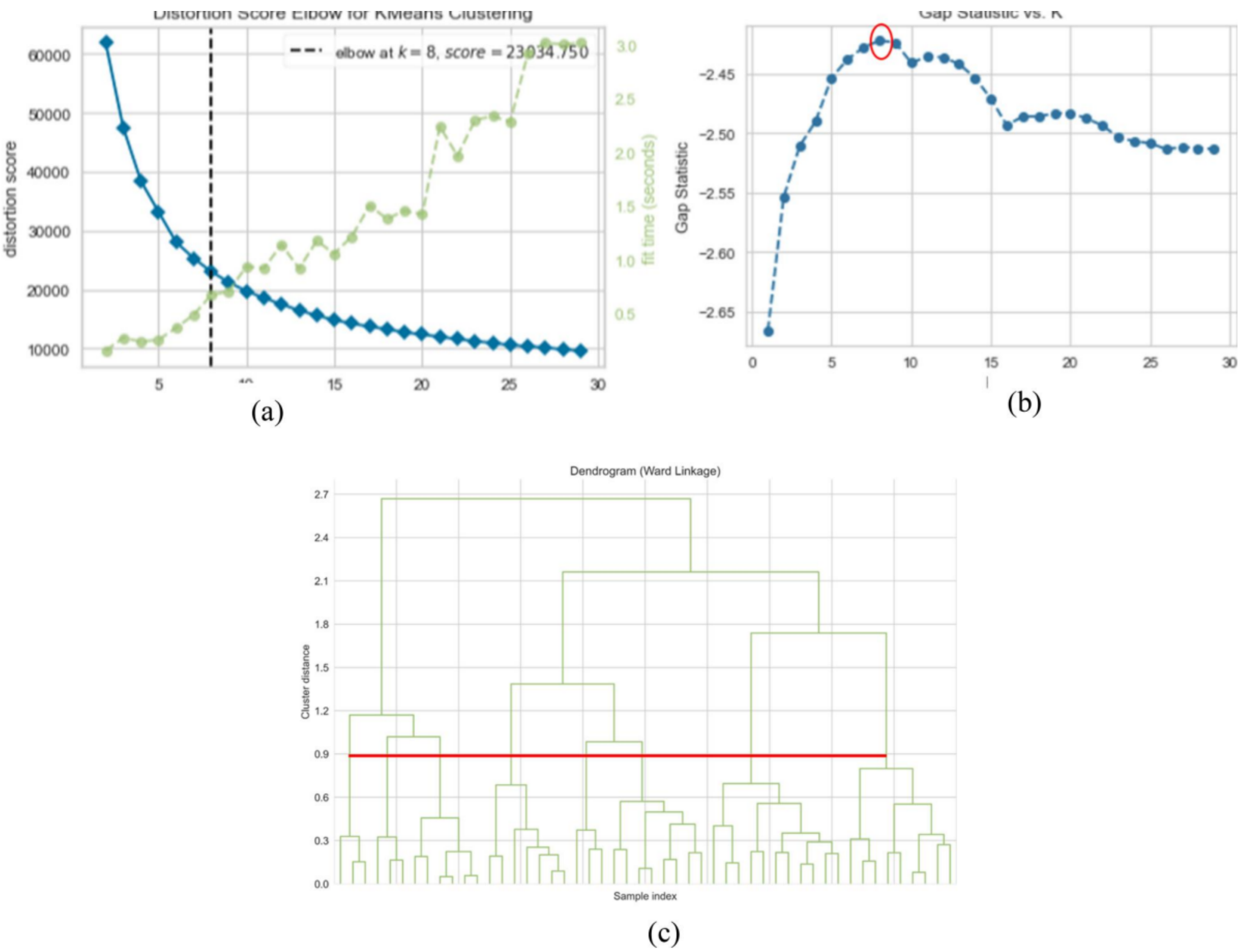


Figure 3. Determination of cluster numbers using elbow (a), gap statistics (b), and Dendrogram.

The dendrogram generated using Ward method. The Ward method in agglomerative hierarchical clustering (AHC) works by initially treating each data point as its own cluster. Throughout the clustering process, it combines clusters that result in the least increase in total within-cluster variance, effectively keeping the clusters as compact as possible. This method uses the squared Euclidean distance to measure the variance within clusters, aiming to minimize this sum as clusters are merged. In the dendrogram produced by the AHC, we used ward technique, the placement of the red line for the cutoff is determined by identifying the largest gap. Gap means distance between the clusters being merged. This is done by examining the vertical distances between successive linkage levels, which represent the dissimilarity measure between clusters being combined. The largest gap method selects the point where the increase in this distance is the most significant. At this point, the clusters below the line are distinctly different from each other, as evidenced by the large jump in cluster distance. Positioning the cutoff at this largest gap ensures that the derived clusters are as homogeneous as possible internally, while the dissimilarity between different clusters is maximized. This method provides a data-driven and objective criterion for deciding the number of clusters, thereby minimizing subjectivity in the clustering process. In our analysis, the red line is placed at the height corresponding to this largest increase in dissimilarity, suggesting a natural division in the dataset and indicating a number of clusters that best captures the inherent structure of the data.

The elbow method is applied by plotting the k-means distortion scores—sums of squared distances within clusters—against a range of k values. The number of clusters is indicated by the "elbow" point, where the graph bends and further increases in k yield diminishing returns in decreasing distortion. This point reflects a balance between complexity and clustering quality. In the provided graph, the elbow is at k = 8, suggesting it as the number of clusters based on the method's criteria for parsimony and sufficient data fit.

4.2.2. Approach 1: k-means and RPCA

As discussed in Subsection 3.2.2, among weather features, precipitation is skewed yet, and RPCA addresses this issue by using a robust estimation method to identify the principal components. It decomposes the data matrix into a low-rank matrix (capturing the main structure of the data) and a sparse matrix (representing the outliers or noise). This decomposition helps to separate the true underlying structure of the data from the effects of the skewed distribution and outliers, leading to a more accurate representation of the principal components. To perform the RPCA, a library, namely "pypropack," is used in Python. The output of RPCA is a NumPy array that keeps 95% variability in the original dataset. Table 5 presents numbers of observations in each cluster.

Table 5. Number of observations in each cluster.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
1,509	2,542	1,707	2,277	2,082	2,433	1,877	3,059

The descriptive analysis of the generated clusters revealed the main attributes of clusters as follows:

1. In weather type 1 most temperatures are distributed from 12 to –5 °C, humidity is between 20% and 50%, wind speed is from 28 to 41 km/h, and precipitation is not observed (o cm).
2. Weather type 2 mainly represents driving conditions in March, April, October, and November. Approximately 90% of the hour's temperature were between 6 and –12 °C, humidity spread from 25% to 50%, one millimeter of precipitation on average presence, and wind blow between 34 and 39 km/h.
3. Weather type 3 represents summer months with 40%, has the highest share. Highest temperature, lowest wind speed, and humidity are found in this cluster. In addition, 94% of hours have zero precipitation.
4. In cluster 4, the temperature is distributed from –1 to 11 °C, humidity spreads from 32% to 65%, precipitation was between 0 and 0.7 mm. Maximum wind speed, which is 35 to 59 km/h, is observed in this weather type.
5. In weather type 5 share of January, February, and December are 27%, 30%, and 28%. Temperature is between –2 and –23 °C, humidity is distributed from 31% to 62%, wind speed is 35 to 52 km/h, and precipitation is between 0.1 and 1.5 mm.
6. Cluster 6 composed of October, November, and March, shows the highest precipitation with temperatures between 10 and –5 °C.
7. Cluster 7 contains temperatures between 2 and –10 °C and humidity between 35% and 60%.

January and February have the highest shares in cluster 8. It has the lowest temperature, highest humidity, and precipitation between 0.3 to millimeters.

4.2.3. Approach 2: Affinity propagation

As mentioned in Subsection 3.4 Affinity Propagation (AP) is a nonparametric clustering method that, unlike K-means, does not require the initial specification of the number of clusters. Upon examining the outcomes, it was found that AP identified nine distinct weather patterns. As Table 6 shows the main drawback of the generated weather types, is the unbalanced number of observations in each cluster. For example, cluster 2 has around 3000 members, while cluster 9 includes fewer than 500 objects. Moreover, some clusters contain a wide range of values. For example, in clusters 4 and 5, temperature values start from –23 and –15 °C, respectively, and reach 30 °C. Similarly, clusters 1, 2, 8, and 9 humidity comprise both lowest and highest values, from 10% to 96%. Furthermore, other than clusters 3, 5, and 6, all clusters have zero precipitation values.

Table 6. Number of observations in each cluster in AP method.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9

1,831	2,993	2,603	1,062	2,886	2,099	1,464	2,129	419

4.2.4. Approach 3: Agglomerative hierarchical clustering

AHC (Agglomerative Hierarchical Clustering) is effective with skewed data due to its bottom-up, hierarchical clustering approach, which exhibits less sensitivity to the distribution of data points. This attribute makes AHC more resilient to outliers and data skewness. However, it is essential to note that AHC might not always be the best choice for all types of skewed data. More specifically, it can suffer drawbacks, such as being computationally expensive for large data sets.

As Figure 3 shows, based on the analysis of the dendrogram, it has been determined that eight clusters provide an appropriate categorization for the various weather conditions.

Analyzing clusters shows that the AHC does not identify weather types accurately. For example, in clusters 1, 4, 6, and 8, the temperature is distributed from –20 to 20 °C, humidity is between 10% and 100%, and wind speed starts from minimum at 10 km/h, to the maximum speed at 59 km/h. Conversely, in clusters 3 and 5, temperature spreads from −15 to 30 °C, precipitation is between 0 and 10 millimeters, and wind speed is 0 to 50 km/h. Table 7 shows the number of observations in each cluster.

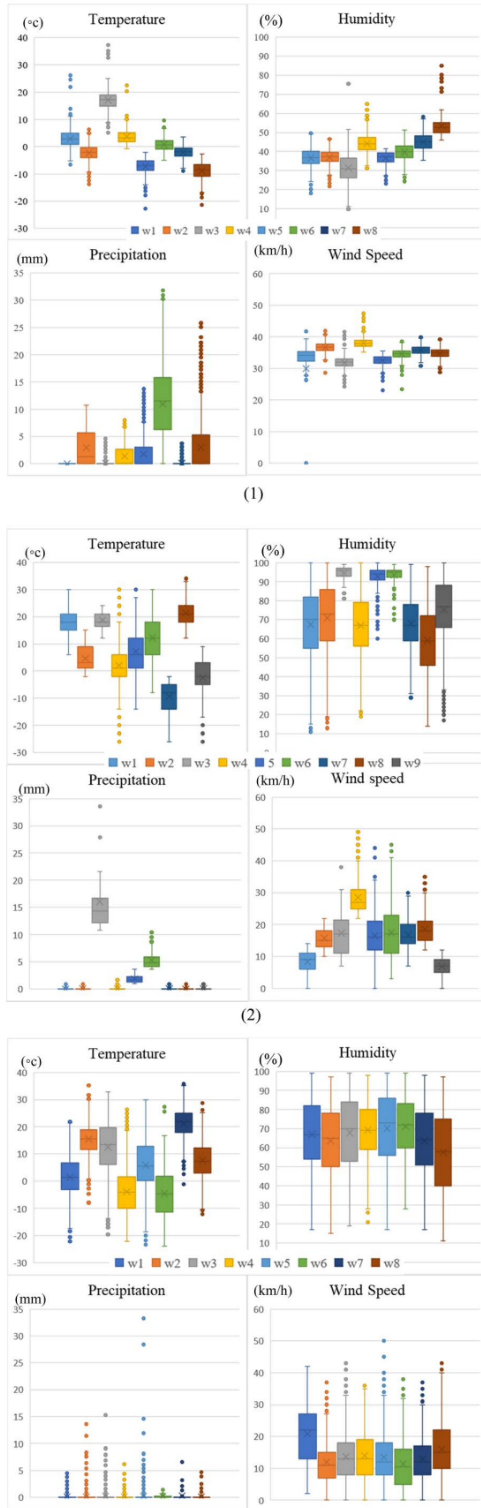
Table 7. Number of observations in each cluster in AHC method.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
2,118	2,661	3,050	2,196	2,354	1,558	2,579	970

4.2.5. Cluster validation

After generating clusters, it is essential to verify the accuracy of the results. One effective method of validating clustering outputs is through content-wise visualization techniques, such as boxplots, which statistically analyze and compare the generated clusters. A boxplot provides a clear and concise way to understand the spread and distribution of data by summarizing the minimum and maximum values, interquartile range (IQR), and potential outliers. In addition, this tool helps gain insight into how parameter values are distributed among different clusters.

Figure 4 compares cluster analysis for three different approaches using boxplots. By examining the distribution of data in various clusters and the range of variations for different clusters, for each weather parameter, it can be concluded that approach 1 provides more accurate weather types.



(3)

Figure 4. Comparing clustering results of approach 1 (1), approach 2 (2), approach 3 (3).

Another approach to validate clustering is to perform a Silhouette analysis. This is a conventional index that literature confirms to be more reliable (Arbelaitz et al., 2013). As Table 8 presents, the results of this technique show that k-means has a higher value, indicating that the first approach better separates weather data. The clustering results suggest that the first approach generates a more effective clustering outcome. Consequently, based on the Silhouette values and content-wise visualization techniques, it can be concluded that the first k-means yielded a more accurate result. Therefore, in the next step of the study, the authors will use the results of this approach.

Table 8. Silhouette values for different clustering approaches.

Clustering techniques	k-means	AHC	AP
Silhouette values	0.54	0.38	0.11

4.3. Sensitivity analysis of weather influences on driving cycle parameters

Clustering methodologies group similar weather conditions, effectively quantifying the variability of driving conditions into a limited number of clusters. For instance, our clustering reveals that Montreal's weather can be categorized into eight distinct groups. This categorization aids in understanding the correlation between changes in driving cycle characteristics and weather types by reducing the need to investigate an infinite number of weather conditions to just eight.

For the sensitivity analysis, we utilized same sample weather data and constructed driving cycles as detailed in Subsection 3.1. Before implementing the MMRM model, we examined the distribution of weather data within each cluster to ensure each had a sufficient sample size. The rationale for using these samples in the sensitivity analysis post-clustering is to compare the R^2 values of MMRM models before and after clustering. This comparison is crucial as R^2 measures the extent to which independent variables can explain variations in dependent variables. An improvement in R^2 suggests that clustering provides a clearer association between weather conditions and driving behaviors. The analysis showed that Cluster 5 had the fewest observations, with a total of 50, whereas Cluster 1 had the most, with 80 observations.

Table 9 compares the R^2 values of models developed for various clustered weather types against a base model, which was formulated using sample data prior to clustering.

Table 9. Comparison R-square between clustered weather types and base.

Cluster	1	2	3	4	5	6	7	8	Base
R^2	0.43	0.53	0.56	0.33	0.53	0.28	0.30	0.35	0.28

As illustrated, except for weather type 6, all other types exhibit higher R^2 values than the base model. This suggests that the model with clustered weather types provides a more accurate representation of the relationship between variations in weather conditions and changes in driving cycle parameters.

Figure 5 presents a heatmap illustrating the results from the developed MMRMs. The cells in the heatmap are categorized into two groups: colorful cells with values, and gray cells. The colorful cells, where the p-values are less than 0.05, indicate statistically significant influences of weather parameters on driving cycle characteristics. The color intensity varies from dark red to dark blue, representing strong positive and negative impacts, respectively, with lighter colors denoting less impact. The coefficient values are displayed within these cells. Conversely, gray cells represent nonsignificant results, suggesting that the corresponding weather parameters do not significantly affect the driving cycle characteristics. In this figure, T stands for temperature, H stands for humidity, W stands for wind speed, and P denotes precipitation. The primary insights drawn from this table include:

1. The influence of weather types on driving cycle parameters is not uniform, likely due to the combined effects of different weather elements.

2. The number of driving cycle parameters which are influenced by weather factors varies across weather types. Specifically, weather type 7 has a lesser impact on driving cycle characteristics compared to other types.

3. Precipitation typically has a significant effect on driving cycle attributes. However, it is absent in weather types 1 and 7, and thus, does not influence the driving cycle parameters in these categories.

4. Temperature has a negative impact on driving cycle attributes for all weather types, with the exception of weather type 3, where its influence is not negative.

5. The effects of humidity and wind speed on driving cycle parameters differ across weather types. As an illustration, humidity positively affects velocity in weather type 2 but negatively in weather type 3.

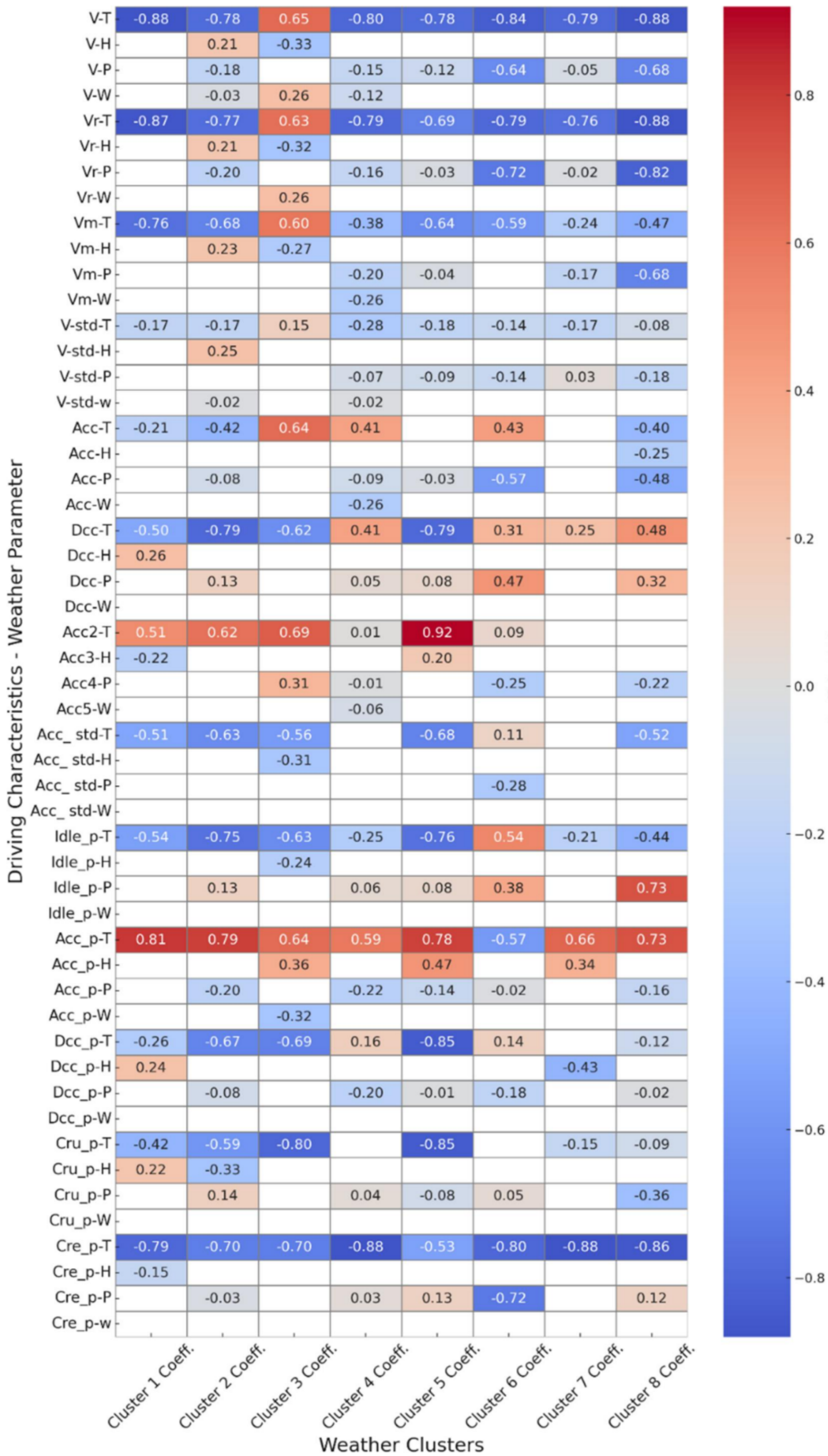


Figure 5. MMRM models output for the developed weather types.

4.4. Driving cycle characteristics analysis

In the next step, driving cycles are constructed for each weather type. In addition, researchers revealed that driving behaviors are different between weekdays and weekends (Agarwal, 2004; Harris & Webber, 2012; Yarahmadi et al., 2024; Zhong et al., 2008). To understand the discrepancies and their connections with weather conditions, separate driving cycles were designed for weekdays and weekends. This resulted in a total of 16 distinct driving cycles: eight for weekdays and eight for weekends. By using this methodology, the dynamics of traffic patterns can be better understood, and the effect of meteorological factors on driving behaviors can be more accurately quantified. The following sections provide more details about different characteristics of the generated driving cycles.

4.4.1. Velocity analysis

Velocity is a crucial driving cycle indicator that significantly impacts both vehicle emissions and fuel consumption. The data indicates that, in general, drivers tend to drive faster on weekends. On weekdays, the highest average speed of 43 km/h is observed under weather condition type 3, while the lowest average speed of 26 km/h is recorded in weather condition type 8. On weekends, the picture is somewhat similar, with the highest average speed of 48 km/h occurring under weather type 3, and the lowest average speed of 28 km/h being observed under weather conditions classified as type 8.

4.4.2. Acceleration analysis

Acceleration and deceleration are key factors that control how vehicles move and stay stable on the road. Vehicle acceleration and deceleration refer to the rates of change in the vehicle's velocity over time, measured in units of meters per second squared (m/s^2) or kilometers per hour per second (kph/s), and represent speed up or slow down, respectively.

Table 10 shows the SAFD of the generated driving cycles. SAFD shows the frequency and magnitude of changes in speed and acceleration during a driving session. An analysis of the SAFD reveals that driving patterns differ between weekdays and weekends. The frequency and magnitude of changes in speed and acceleration vary depending on the day of the week. In addition, in all-weather types except weather type 4, accelerations are distributed from 0 to more than 0.8 meters per second squared (m/s^2).

Table 10. Comparison between SAFD of weekdays and weekend.

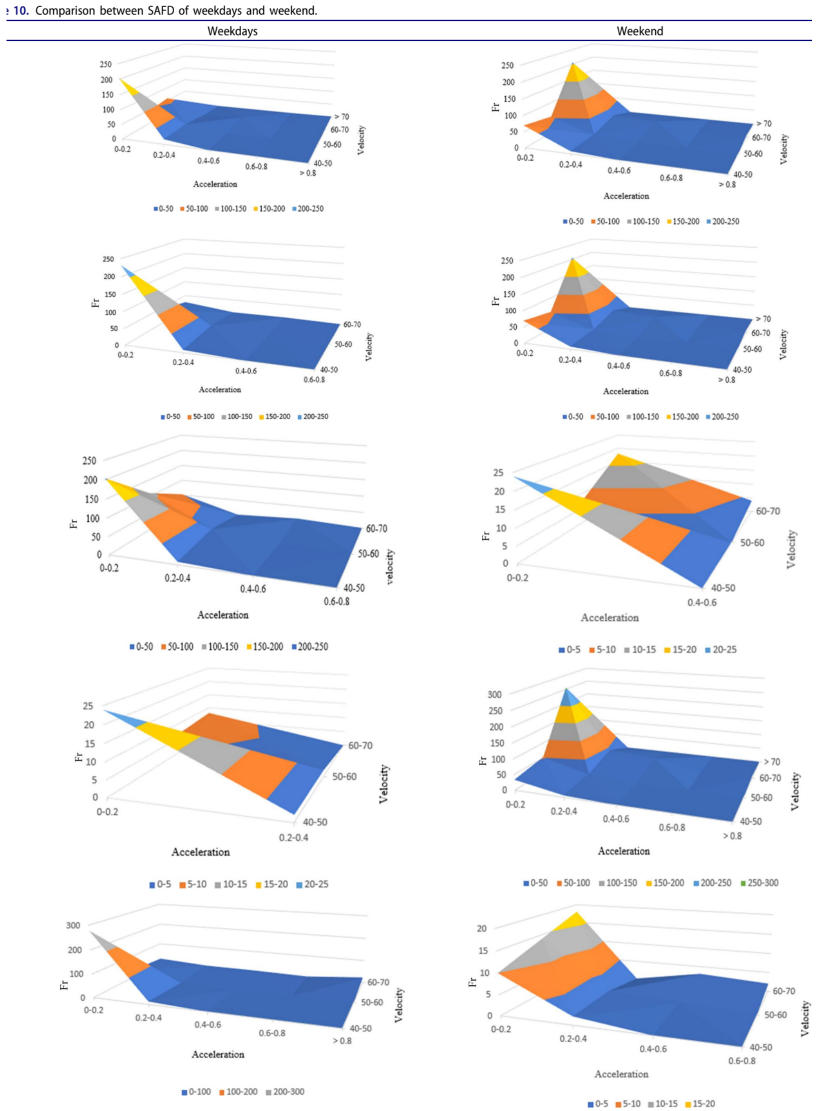
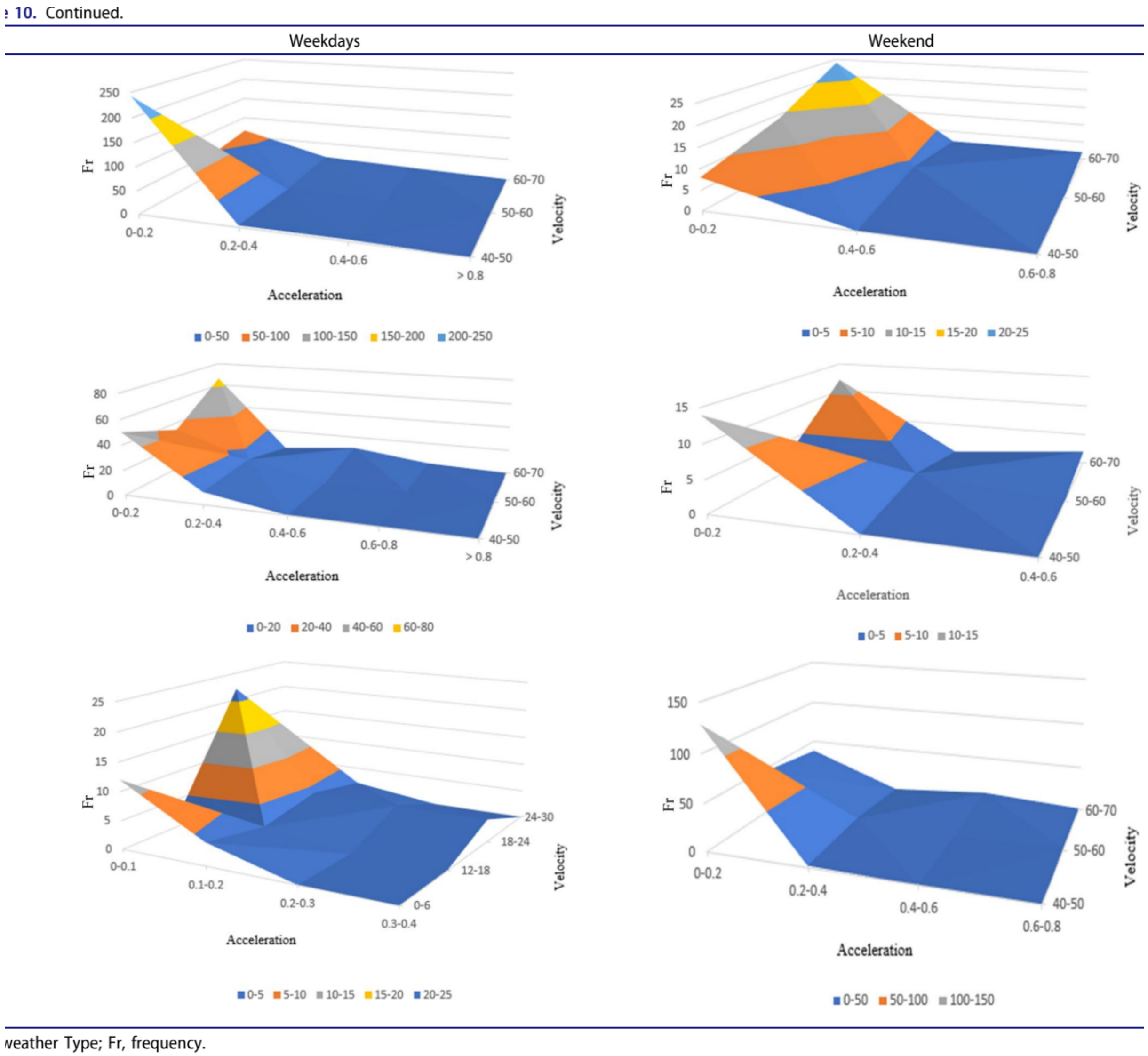


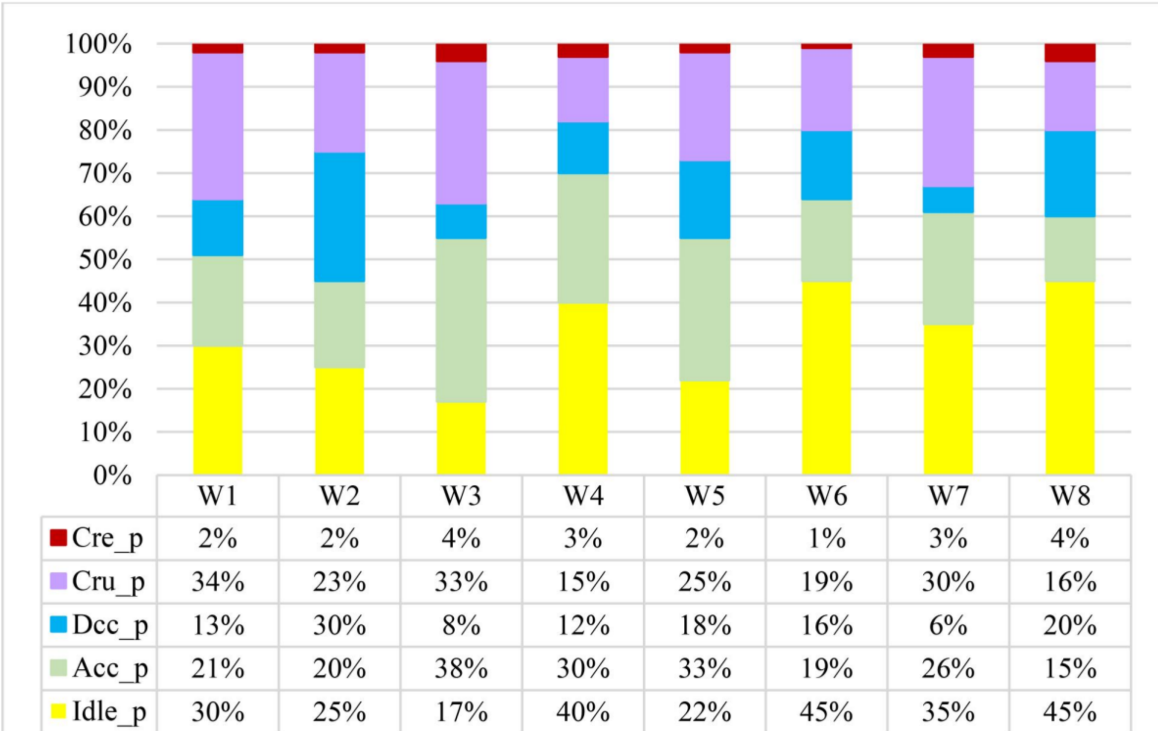
Table 10 (continued). Comparison between SAFD of weekdays and weekend.



WT, weather Type; Fr, frequency.

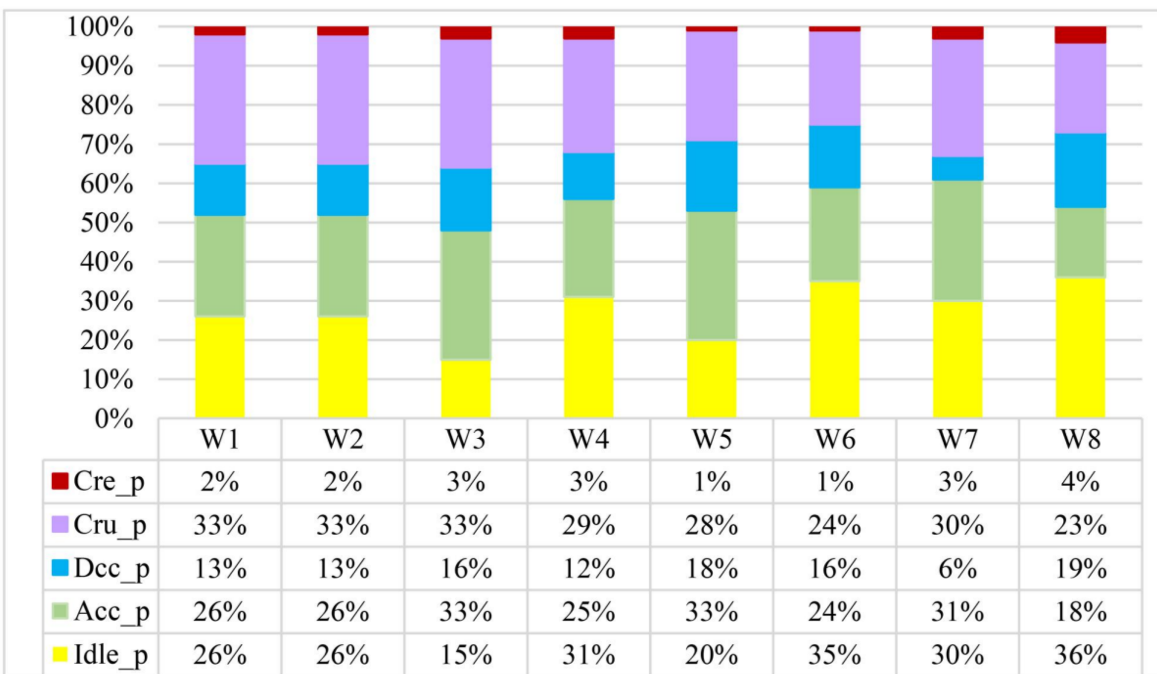
4.4.3. Share of driving modes

Figure 6 portrays the distribution of various driving modes across weekdays and weekends under different weather conditions. Analyzing the mode distribution on different days, except for weather types 1 and 2 on weekends, reveals that they are dissimilar, indicating varying driving patterns in diverse driving scenarios. During weekdays, the time proportion of acceleration values exhibit oscillations ranging from 15% to 38%, while on weekends, they range from 18% to 33%. Additionally, the average deceleration values range between 6% to 30% on weekdays and 6% to 19% on weekends. In nearly all cases, except for weather type 8 on weekends, acceleration percentages exceed deceleration percentages. The idling time percentages vary on weekdays, with weather types 3 exhibiting 17%, while weather type 8 has almost 45%. The share of cruising mode is from 15 to 34 on weekdays, but on weekends, it ranges from 23% to 33%.



	W1	W2	W3	W4	W5	W6	W7	W8
Cre_p	2%	2%	4%	3%	2%	1%	3%	4%
Cru_p	34%	23%	33%	15%	25%	19%	30%	16%
Dcc_p	13%	30%	8%	12%	18%	16%	6%	20%
Acc_p	21%	20%	38%	30%	33%	19%	26%	15%
Idle_p	30%	25%	17%	40%	22%	45%	35%	45%

(a)



	W1	W2	W3	W4	W5	W6	W7	W8
Cre_p	2%	2%	3%	3%	1%	1%	3%	4%
Cru_p	33%	33%	33%	29%	28%	24%	30%	23%
Dcc_p	13%	13%	16%	12%	18%	16%	6%	19%
Acc_p	26%	26%	33%	25%	33%	24%	31%	18%
Idle_p	26%	26%	15%	31%	20%	35%	30%	36%

(b)

Figure 6. Share of modes for all weather types on weekdays (a) and weekends (b).

4.5. Comparative analysis of newly constructed driving cycles versus established models

After creating the driving cycles, the first step involved comparing these newly constructed cycles with existing ones, such as the Montreal Local Driving Cycle (Nouri & Morency, 2017) and the Standard Driving Cycle (SDC) from Canada's 5-Cycle test. To illustrate their differences, as presented in Table 11 the authors used descriptive statistics of the constructed driving cycle parameters' values alongside those from previous studies. This comparison included the minimum and maximum values of driving cycle parameters on weekdays and weekends. The results confirmed that the values of the former driving cycles mostly do not fall within the range of the newly generated driving cycle values. For instance, the minimum and maximum velocities were 26 and 48 km/h, respectively, compared to 25 and 34 km/h in the previous local and standard driving cycles.

Additionally, a one sample t-test was performed to determine whether the mean values of the new driving cycle parameters significantly differed from those in the previous local and standard cycles. The null hypothesis posited that the new driving cycle's mean value, μ, for each parameter, would equal the corresponding values in the local and standard driving cycles (μ_0). Conversely, the alternative hypothesis suggested that the driving cycle's mean value for each parameter would not equal the corresponding ones in the local and standard driving cycles. The hypotheses are formulated as follows:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

To perform this test, the study utilized the SciPy library in Python. Equation (1) shows t-test statistic formula:

$$t = (\mu - \mu_0) / (s / \sqrt{n}) \quad (1)$$

where standard deviation as s and the number of newly constructed driving cycles as n.

Table 11 presents the results of the t-test, indicating significant differences between the generated driving cycles and the two former ones. The table clearly shows that the p-values of the driving cycle parameters are less than 0.05, suggesting that the proposed approach significantly differs from the preexisting ones.

Table 11. Differences between the developed driving cycles 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	26	26	73	2	0	–0.6	0	0.2	17%	15%	6%	15%	1%
	Max-WDs	43	43	91	6	0.8	0	1.5	1.2	45%	38%	30%	34%	4%
	Min-WE	28	28	73	2	0.1	–0.2	1	0	15%	18%	6%	23%	1%
	Max-WE	48	48	92	10	0.3	–0.1	0.3	0.6	36%	33%	19%	33%	4%
	μ	37.68	36	79.37		0.19	–0.15		0.51	29.87%	25%	16%	26.6%	2.5%
	s	6.3	6.9	5.84		0.17	0.13		0.41	9.13%	7.3%	6.4%	5.8%	1.03%
(Nouri, 2015) (μ_0)		25	32	104	–	0.43	–0.5	–	0.29	19.9%	34.5%	12.5%	31.2%	1.6%
5-cycle test (city) (μ_0)		34	–	90	–	5.3	–	–	–	18%	–	–	–	–
One sample T-test (Nouri, 2015)	T-statistic	7.9	2.3	–	–	–	–	–	2.1	4.3	–5.2	2.21	–3.18	3.2
	p-value	0.0	0.04	–	–	0	–	–	0.04	0	0	0.04	0	0
One sample T-test 5-cycle test (city)	T-statistic	2.32	–	–7.2	–	–110	–	–	–	5.1	–	–	–	–
	p-value	0.03	–	0	–	0	–	–	–	0	–	–	–	–

This highlights that a single driving cycle cannot adequately represent driving behaviors under varying conditions, emphasizing the importance of considering the influence of weather when constructing driving cycles.

4.6. Determining the optimal sets of driving cycles

Descriptive analysis revealed that the proportion of different modes of transport varies with weather types on weekends and weekdays, indicating that the driving cycles differ. To determine optimal number of driving cycles we need to measure similarity of generated driving cycles which represent driving behaviors in different driving conditions. Measuring similarity is important because similar driving cycles can show similar driving behaviors; therefore, they should be integrated and synthesize a new driving cycle to represent driving behaviors for the relevant weather conditions. To do that, we utilized a similarity matrix which is adept at facilitating the pairwise comparison of the multidimensional arrays representing driving cycles. The construction of the similarity matrix is a nuanced process that involves the normalization of parameter data to a common scale to ensure comparability and then employing a pairwise Euclidean distance computation across all possible pairs of driving cycles. The similarity between driving cycles is quantified with a value between 0 and 1, which in this case, is derived from calculating the Euclidean distance between vectors of the 13 parameters for each pair of driving cycles. Euclidean distance between two points is calculated using Equation (2).

$$d(P, Q) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \quad (2)$$

where p_i and q_i are the values of the i^{th} parameter for driving cycles P and Q, respectively, and n is the total number of parameters. After calculating the Euclidean distances, these are then transformed into similarity scores. This transformation is typically achieved by inverting the distances so that a smaller distance (indicating higher similarity) results in a higher similarity score.

One common method is to use the formula $1/(1 + d(P, Q))$ which bounds the similarity score between 0 and 1. The closer the score is to 1, the more similar the driving cycles are with respect to the parameters considered.

As is presented in Figure 7, the similarity matrix visually encodes the level of resemblance, with darker shades representing higher values and thus greater similarity. The findings from this matrix are pivotal, as they demonstrate that weekends in weather types 1 and 2 exhibit a 0.92 similarity, which is a substantial overlap, suggesting that they could be effectively classified as a single driving cycle for the purposes of modeling and analysis. This high degree of similarity may reflect a commonality in driving behavior patterns during these periods, perhaps due to reduced variability in driving conditions or consistent driver responses to these conditions. The next level of similarity is 0.32, which means that the compared driving cycles shows the driving behaviors are 68% are different which is considerable difference. Therefore, it is inferred that the optimal number of driving cycles to accurately portray driving behaviors in all driving situations for the chosen road category is 15.

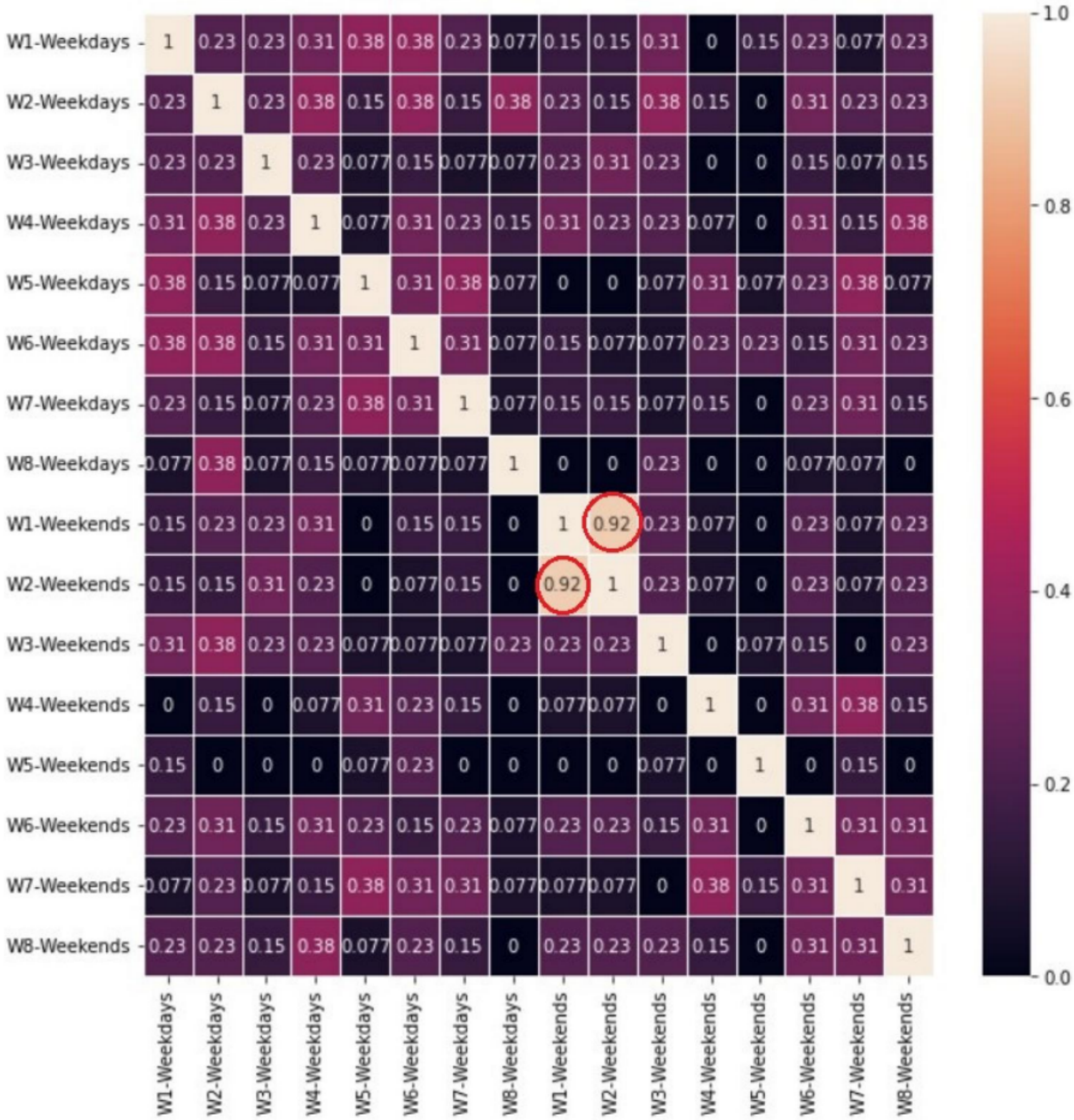


Figure 7. Similarity matrix of the generated driving cycles in different weather conditions.

5. Conclusion

This study mainly targets three concerns that are associated with the volatility of the driving cycle under various weather circumstances, utilizing scientific investigation and analysis. (1) limitations in representing real-world driving conditions with current driving cycle development methodologies; (2) insufficient analysis of the impact of weather characteristics on driving cycle parameters; (3) scarcity of discussions on a systematic approach to identify different weather types in vehicle emissions context.

First, an optimal number of driving cycles is determined to represent the driving patterns in various driving conditions. The proposed model effectively identified 15 driving cycles that represent driving behaviors under various conditions, but it's important to note that these findings are currently specific to a certain type of segments. Further validation is necessary to determine if these results can be generalized to other segment types. Of the 15 driving cycles generated in this study, 8 represent weekday driving behaviors and 7 represent weekend driving behavior. The aforementioned finding underscores the crucial need to incorporate weather conditions, weekday and weekend driving behaviors in developing fuel consumption testing protocols. Doing so is essential to ensure accurate guidance for the automobile industry and effective decarbonization planning.

Second, a MMRM approach is developed to identify the most essential weather variables playing a significant role in the variability of driving behaviors. The model revealed that temperature exerts a more significant influence on deceleration compared to other variables, whereas precipitation impacts velocity, wind speed affects acceleration, and humidity displays minimal effects on driving cycle parameters. To enhance their understanding of the influence of these factors on driving parameters, the researchers opted to isolate the effects of weather conditions by selecting a road type prior to performing the regression analysis. The model's results shed new light on the association between weather conditions and driving cycles, offering insights that may prove useful for researchers exploring the impact of climate change on driving behavior.

Lastly, an unsupervised machine learning method was employed to cluster weather conditions based on factors affecting driving cycle parameters. The resulting clusters revealed the existence of 8 distinct weather types in the study area. Our analysis for approach 1 demonstrates that wind speed conditions across the eight weather clusters appear quite similar. Initially, this observation might suggest a limited differentiation capacity of wind speed within our clustering model. However, it is imperative to consider the multidimensional integration of weather parameters—precipitation, wind speed, humidity, and temperature—which collectively define our clustering criteria. The inclusion of wind speed, despite its similar statistical distribution across clusters, plays a crucial role in the holistic weather profiling. This is because wind speed can interact variably with other weather parameters, influencing overall weather conditions and subsequently, driving behaviors. Furthermore, the inclusion of humidity in the weather type categorization, despite its minimal impact on driving cycle parameters, was based on comprehensive multivariate modeling which considered interactions between various weather conditions. While the results indicate that humidity has a lesser influence on driving behaviors compared to other weather parameters such as temperature, wind speed, and precipitation, it still contributes to the overall atmospheric conditions that could subtly affect driving dynamics.

A major constraint of this study is the limited availability of data. Numerous factors can characterize weather conditions, yet the prevailing weather data compiler, Environment Canada, presently only collects a restricted set of weather parameters on an hourly basis. Notably, visibility is not included in the hourly data, snow and rain are reported collectively, and wind direction has significant values of missing values. These parameters can affect driving behavior, vehicle traction, and ultimately, emissions. Recognizing their impacts, future research could aim to disaggregate these precipitation types if more detailed meteorological data becomes available. This would allow for a more nuanced understanding and modeling of the specific impacts of driving conditions on driving cycles.

The following points are suggested for future research:

- There is mounting evidence that weather variables are expected to undergo significant changes in the future as a result of climate change. To better inform mitigation strategies, decision-makers may benefit from a model that can predict how changes in weather parameters will affect driving cycle parameters. In addition, such a model could help ensure that legislative and regulatory measures are updated in a more informed and proactive manner.
- This study utilizes taxi data to generate driving cycles, and applying the developed model to passenger cars can potentially improve the accuracy of driving cycle predictions.
- Given that Montreal has a continental climate, extending the proposed methodology to other urban areas with different climate types can enhance the comprehensiveness of this study.

Validation of the proposed model across different types of segments. This would not only test the robustness of the model but also potentially broaden its applicability in understanding driving behaviors under a variety of driving conditions.

Authors' contributions

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.

Disclosure statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Funding

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

ORCID

Asad Yarahmadi — 0000-0001-9640-4023

Catherine Morency — 0000-0002-3211-4243

Martin Trépanier — 0000-0001-8408-8035

References

Abdi Kordani, A., Rahmani, O., Abdollahzadeh Nasiri, A. S., & Boroomandrad, S. M. (2018). Effect of adverse weather conditions on vehicle braking distance of highways. *Civil Engineering Journal, 4*(1), 46. https://doi.org/10.28991/cej-030967

Abele, L., & Møller, M. (2011). The relationship between road design and driving behavior. In *RSS 2011: Road Safety and Simulation 2011 Conference*.

Achour, H., & Olabi, A. G. (2016). Driving cycle developments and their impacts on energy consumption of transportation. *Journal of Cleaner Production, 112*, 1778–1788. https://doi.org/10.1016/j.jclepro.2015.08.007

Agarwal, A. (2004). *A comparison of weekend and weekday travel behavior characteristics in urban areas*. University of South Florida. https://digitalcommons.usf.edu/etd

Al-Samari, A. (2017). Real-world driving cycle: Case study of Baqubah, Iraq. *Diyala Journal of Engineering Sciences, 10*(2), 39–47.

Ameknassi, L., Ait-Kadi, D., & Keivanpour, S. (2016). Incorporating design for environment into product development process: an integrated approach. *IFAC-PapersOnLine, 49*(12), 1460–1465. https://doi.org/10.1016/j.ifacol.2016.07.777

Amirjamshidi, G., & Roorda, M. J. (2013). Development of simulated driving cycles: Case study of the Toronto waterfront area: Enhanced reader. *Transportation Research Board.* No. 13–2648.

Andrey, J., Hambly, D., Mills, B., & Afrin, S. (2013). Insights into driver adaptation to inclement weather in Canada. *Journal of Transport Geography, 28*, 192–203. https://doi.org/10.1016/j.jtrangeo.2012.08.014

Anida, I. N., Ismail, I. S., Norbakyah, J. S., Atiq, W. H., & Salisa, A. R. (2017). Characterisation and development of driving cycle for work route in Kuala Terengganu. *International Journal of Automotive and Mechanical Engineering, 14*(3), 4508–4517. https://doi.org/10.15282/ijame.14.3.2017.9.0356

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition, 46*(1), 243–256. https://doi.org/10.1016/j.patcog.2012.07.021

Ashtari, A., Bibeau, E., & Shahidinejad, S. (2014). Using large driving record samples and a stochastic approach for real-world driving cycle construction: Winnipeg driving cycle. *Transportation Science, 48*(2), 170–183. https://doi.org/10.1287/trsc.1120.0447

Bodisco, T., & Zare, A. (2019). Practicalities and driving dynamics of a real driving emissions (RDE) Euro 6 regulation homologation test. *Energies, 12*(12), 2306. https://doi.org/10.3390/en12122306

Candès, E. J., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis. *Journal of the ACM (JACM), 58*(3), 1–37. https://doi.org/10.1145/1970392.1970395

Chauhan, B. P., Joshi, G. J., & Parida, P. (2020). Development of candidate driving cycles for an urban arterial corridor of Vadodara City. *International Journal of Transport Economics, Engineering and Law, 4*(75), 16.

Chen, C., Zhao, X., Liu, H., Ren, G., & Liu, X. (2019). Influence of adverse weather on drivers' perceived risk during car following based on driving simulations. *Journal of Modern Transportation, 27*(4), 282–292. https://doi.org/10.1007/s40534-019-00197-4

Chen, T., Zhang, M., & Wei, L. (2014). Driver behavior on combination of vertical and horizontal curves of mountainous freeways. *Mathematical Problems in Engineering, 2014*, 1–9. https://doi.org/10.1155/2014/432841

Chowdhury, N. F. A. (2015). Ambient temperature effects on driving. *Procedia Manufacturing, 3*, 3123–3127. https://doi.org/10.1016/j.promfg.2015.07.860

Christian, T., & Jensen, T. C. (2014). General rights weather and road capacity. *Danish Journal of Transportation Research*, 1603–9696, 19.

Cui, Y., Zou, F., Xu, H., Chen, Z., & Gong, K. (2022). A novel optimization-based method to develop representative driving cycle in various driving conditions. *Energy, 247*, 123455. https://doi.org/10.1016/j.energy.2022.123455

Degraeuwe, B., & Weiss, M. (2017). Does the new european driving cycle (NEDC) really fail to capture the NOX emissions of diesel cars in Europe? *Environmental Pollution (Barking, Essex: 1987), 222*, 234–241. https://doi.org/10.1016/j.envpol.2016.12.050

Eckert, J. J., Santiciolli, F. M., Bertoti, E., Costa, E. d S., Corrêa, F. C., Silva, L. C. d A. e., & Dedini, F. G. (2018). Gear shifting multi-objective optimization to improve vehicle performance, fuel consumption, and engine emissions. *Mechanics Based Design of Structures and Machines, 46*(2), 238–253. https://doi.org/10.1080/15397734.2017.1330156

Elsagheer Mohamed, S. A., Alshalfan, K. A., Al-Hagery, M. A., & Ben Othman, M. T. (2022). Safe driving distance and speed for collision avoidance in connected vehicles. *Sensors, 22*(18), 7051. https://doi.org/10.3390/s22187051

EPA. (2014). *Evaporative Emissions from On-Road Vehicles in MOVES2014* (EPA-420-R-14-014, September 2014).

Esteves-Booth, A., Muneer, T., Kirby, H., Kubie, J., & Hunter, J. (2001). The measurement of vehicular driving cycle within the city of Edinburgh. *Transportation Research Part D: Transport and Environment, 6*(3), 209–220.

European Commission Joint Research Center. (2013). *Regulated emissions of a Euro 5 passenger car measured over different driving cycles*.

Faria, M. V., Baptista, P. C., & Farias, T. L. (2017). Main. *Transportation Research Procedia, 27*, 953–960. https://doi.org/10.1016/j.trpro.2017.12.038

Fotouhi, A., & Montazeri-Gh, M. (2012). Tehran driving cycle development using the k-means clustering method. *Scientia Iranica, 20*(2), 286–293.

Galgamuwa, U., Perera, L., & Bandara, S. (2016). A representative driving cycle for the southern expressway compared to existing driving cycles. *Transportation in Developing Economies, 2*(2), 22. https://doi.org/10.1007/s40890-016-0027-4

Gebisa, A., Gebresenbet, G., Gopal, R., & Nallamothu, R. B. (2021). Driving cycles for estimating vehicle emission levels and energy consumption. *Future Transportation, 1*(3), 615–638. https://doi.org/10.3390/futuretransp1030033

Greene, D. L., Khattak, A. J., Liu, J., Wang, X., Hopson, J. L., & Goeltz, R. (2017). What is the evidence concerning the gap between on-road and environmental protection agency fuel economy ratings? *Transport Policy, 53*, 146–160. https://doi.org/10.1016/j.tranpol.2016.10.002

Günther, R., Wenzel, T., Wegner, M., & Rettig, R. (2017). Big data driven dynamic driving cycle development for busses in urban public transportation. *Transportation Research Part D: Transport and Environment, 51*, 276–289. https://doi.org/10.1016/j.trd.2017.01.009

Harris, C. B., & Webber, M. E. (2012). A temporal assessment of vehicle use patterns and their impact on the provision of vehicle-to-grid services. *Environmental Research Letters, 7*(3), 034033. https://doi.org/10.1088/1748-9326/7/3/034033

Jägerbrand, A. K., & Sjobergh, J. (2016). Effects of weather conditions, light conditions, and road lighting on vehicle speed. *SpringerPlus, 5*(1), 505. https://doi.org/10.1186/s40064-016-2124-6

Kaymaz, H., Korkmaz, H., & Erdal, H. (2019). Development of a driving cycle for istanbul bus rapid transit based on real-world data using stratified sampling method. *Transportation Research Part D: Transport and Environment, 75*, 123–135. https://doi.org/10.1016/j.trd.2019.08.023

Kazemi Miyangaskari, M., Mehrregan, M. R., Safari, H., Keivanpour, S., & Dehghan Nayer, M. (2023). 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). *Industrial Management Studies, 21*(69), 1–42. https://doi.org/10.22054/jims.2023.70278.2815

Kazemi Miyangaskary, M., Keivanpour, S., & Lamghari, A. (2023). Optimizing closed-loop supply chain in the electric vehicle battery industry: A fully fuzzy approach. *IFIP International Conference on Product Lifecycle Management.* Switzerland: Springer Nature Switzerland.

Keivanpour, S. (2022). Sustainability balanced scorecard approach to internet of things enabled logistics systems. *Engineering Management Journal, 34*(3), 450–474. https://doi.org/10.1080/10429247.2021.1946320

Kilpeläinen, M., & Summala, H. (2007). Effects of weather and weather forecasts on driver behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour, 10*(4), 288–299. https://doi.org/10.1016/j.trf.2006.11.002

Ladubec, C., & Bensebaa, F. (2021). *Assessment of energy efficiency ratios of electric and fuel cell light duty vehicles in Canada*. Final Report.

Lewandowski, D., Cooke, R. M., & Tebbens, R. J. D. (2007). Sample-based estimation of correlation ratio with polynomial approximation. *ACM Transactions on Modeling and Computer Simulation, 18*(1), 1–17. https://doi.org/10.1145/1315575.1315578

Liu, H., Rodgers, M. O., Liu, F., & Guensler, R. (2020). Bayesian approach in estimating the road grade impact on vehicle speed and acceleration on freeways. *Transportmetrica A: Transport Science, 16*(3), 602–625. https://doi.org/10.1080/23249935.2020.1722280

Ma, R., He, X., Zheng, Y., Zhou, B., Lu, S., & Wu, Y. (2019). Real-world driving cycles and energy consumption informed by large-sized vehicle trajectory data. *Journal of Cleaner Production, 223*, 564–574. https://doi.org/10.1016/j.jclepro.2019.03.002

Muhammad, J. (2022). Development of driving cycle for passenger cars and estimation of vehicular exhaust emission factors. *Journal of Environmental Pollution and Control, 5*(1), 1–11. https://doi.org/10.15744/2639-9288.5.103

Nouri, P. (2015). *Enhancing the gasoline vehicles' CO2 emissions estimation in montreal* (PhD thesis). École Polytechnique de Montréal, PolyPublie. https://publications.polymtl.ca/2038

Nouri, P., & Morency, C. (2017). Evaluating microtrip definitions for developing driving cycles. *Transportation Research Record: Journal of the Transportation Research Board, 2627*(1), 86–92. https://doi.org/10.3141/2627-10

Pouresmaeili, M. A., Aghayan, I., & Taghizadeh, S. A. (2018). Development of mashhad driving cycle for passenger car to model vehicle exhaust emissions calibrated using on-board measurements. *Sustainable Cities and Society, 36*, 12–20. https://doi.org/10.1016/j.scs.2017.09.034

Ragione, L. D., & Giovanni, M. (2016). The influence of road gradient in an integrated approach of real driving cycles and emissions factors model. In *Transportation Research Procedia* (pp. 3179–3188). Elsevier B.V. https://doi.org/10.1016/j.trpro.2016.05.258

Ratner, B. (2009). The correlation coefficient: Its values range between 1/1, or do they. *Journal of Targeting, Measurement and Analysis for Marketing, 17*(2), 139–142. https://doi.org/10.1057/jt.2009.5

Ribeiro, S. M. (2021). *Imputation by decomposition and by time series pattern: Novel imputation methods for missing data in time series* (Master's thesis). Universidade Federal de Minas Gerais, Repositório Institucional da UFMG. http://hdl.handle.net/1843/46099

Rios-Torres, J., Liu, J., & Khattak, A. (2019). Fuel consumption for various driving styles in conventional and hybrid electric vehicles: Integrating driving cycle predictions with fuel consumption optimization. *International Journal of Sustainable Transportation, 13*(2), 123–137. https://doi.org/10.1080/15568318.2018.1445321

Roy, F., & Morency, C. (2020). Comparing driving cycle development methods based on Markov chains. *Transportation Research Record: Journal of the Transportation Research Board, 2675*(3), 212–221. https://doi.org/10.1177/0361198120968829

Seers, P., Nachin, G., & Glaus, M. (2015). Development of two driving cycles for utility vehicles. *Transportation Research Part D: Transport and Environment, 41*, 377–385. https://doi.org/10.1016/j.trd.2015.10.013

United Nations. (2014). *Addendum 15: Global Technical Regulation No. 15 worldwide harmonized light vehicles test procedure global technical regulation on worldwide harmonized light vehicles test procedure contents*.

Wood, A. L. (2015). *A climatology of North American air masses and their extremes in Montreal, Quebec*. McGill University.

Wu, Y., Huang, Z., Zheng, Y., Liu, Y., Li, H., Che, Y., Peng, J., & Teodorescu, R. (2023). Spatial–temporal data-driven full driving cycle prediction for optimal energy management of battery/supercapacitor electric vehicles. *Energy Conversion and Management, 277*, 116619. https://doi.org/10.1016/j.enconman.2022.116619

Xu, S., Qiao, X., Zhu, L., Zhang, Y., Xue, C., & Li, L. (2016). Reviews on determining the number of clusters. *Applied Mathematics & Information Sciences, 10*(4), 1493–1512. https://doi.org/10.18576/amis/100428

Yarahmadi, A. (2023). *Polytechnique Montréal a methodological framework to develop optimal sets of driving cycles for a region* (PhD thesis). Polytechnique Montréal, PolyPublie. https://publications.polymtl.ca/57092/

Yarahmadi, A., Morency, C., & Trepanier, M. (2023). New data-driven approach to generate typologies of road segments. *Transportmetrica A: Transport Science, 20*(2), 1–24. https://doi.org/10.1080/23249935.2022.2163206

Yarahmadi, A., Morency, C., & Trépanier, M. (2024). A context-aware approach to construct driving cycles in a volatile urban environment. *SSRN.* https://doi.org/10.2139/ssrn.4856910

Yeo, J., Lee, J., & Jang, K. (2021). The effects of rainfall on driving behaviors based on driving volatility. *International Journal of Sustainable Transportation, 15*(6), 435–443. https://doi.org/10.1080/15568318.2020.1756543

Zahid, M., Habib, M. F., Ijaz, M., Ameer, I., Ullah, I., Ahmed, T., & He, Z. (2024). Factors affecting injury severity in motorcycle crashes: Different age groups analysis using Catboost and SHAP techniques. *Traffic Injury Prevention, 25*(3), 472–481. https://doi.org/10.1080/15389588.2023.2297168

Zhai, Y., et al. (2018). A context-aware evaluation method of driving behavior. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (pp. 462–474). Springer Verlag.

Zhang, J., Song, G., Gong, D., Gao, Y., Yu, L., & Guo, J. (2018). Analysis of rainfall effects on road travel speed in Beijing, China. *IET Intelligent Transport Systems, 12*(2), 93–102. https://doi.org/10.1049/iet-its.2017.0039

Zhong, M., Hunt, J. D., & Lu, X. (2008). Studying differences of household weekday and weekend activities a duration perspective. *Transportation Research Record: Journal of the Transportation Research Board, 2054*(1), 28–36. https://doi.org/10.3141/2054-04