Clustering and Disjoint Principal Component Analysis of emissions and driving volatility data collected from a hybrid electric vehicle in real drive conditions

Paulo Fernandes, PhD.
Junior Researcher, Mechanical Engineering,
University of Aveiro, Centre for Mechanical Technology and Automation (TEMA),
Campus Universitário de Santiago, 3810-193 Aveiro - Portugal,
E-mail: paulo.fernandes@ua.pt
ORCID http://orcid.org/0000-0001-5448-5048

Eloisa Macedo, PhD.
Junior Researcher, Mechanical Engineering,
University of Aveiro, Centre for Mechanical Technology and Automation (TEMA),
Campus Universitário de Santiago, 3810-193 Aveiro - Portugal
E-mail: macedo@ua.pt
ORCID http://orcid.org/0000-0003-1503-1718

Ricardo F. Tomás, MSc.
PhD Student, Mechanical Engineering,
University of Aveiro, Centre for Mechanical Technology and Automation (TEMA),
Campus Universitário de Santiago, 3810-193 Aveiro - Portugal
E-mail: ricardotomas@ua.pt
ORCID https://orcid.org/0000-0001-5256-1721

Margarida C. Coelho, PhD.
Assistant Professor with Habilitation, Mechanical Engineering
University of Aveiro, Dept. Mechanical Engineering / Centre for Mechanical Technology and Automation (TEMA), Campus Universitário de Santiago, 3810-193 Aveiro - Portugal
E-mail: margarida.coelho@ua.pt
ORCID: https://orcid.org/0000-0003-3312-191X

ABSTRACT
Despite the fuel use and emission benefits of Hybrid Electric Vehicles (HEVs), few studies have characterized in detail emission patterns and driving volatility profiles from HEVs in different road types under Real Driving Emission (RDE) conditions. This paper characterized second-by-second tailpipe emissions, vehicle engine, and dynamics from a 2020 Toyota HEV sub-compact on a 44 km driving route over rural, urban, and highway roads in the Aveiro region (Portugal). Driving volatility was represented by six driving styles based on combinations of acceleration/deceleration and vehicular jerk (the rate at which an object's acceleration changes with respect to the time). Clustering and Disjoint Principal Component Analysis (CDPCA) was applied to examine the relationships between emissions, engine, internal combustion engine (ICE) status, roadway characteristics, and vehicular jerk types. Although the urban route yielded lower carbon dioxide and nitrogen oxides emissions than rural and highway routes did, it resulted in highly volatile driving behaviors at low speeds (< 45 $km.h^{-1}$). Both route type and HEV ICE operating behavior showed to have an impact on the distribution of vehicular jerk types. CDPCA constrained to road sector exhibited different shapes in the clusters of the jerk types between ICE operation status. This paper can provide insights into RDE analysis of the new generation of HEVs about the characterization of volatile driving behaviors. Such information can be integrated into vehicle electronic car units and navigation systems to provide feedback for drivers about their driving behavior in terms of high emission rates and jerkings to the vehicle.

Keywords: Hybrid Electric Vehicles, Tailpipe Emissions, Vehicular Jerk Classification, Clustering and Disjoint Principal Component Analysis.

1. INTRODUCTION AND LITERATURE REVIEW

Hybrid electric vehicles (HEVs) are getting popular in the United States of America (USA) and European Union (EU) countries because of their notable fuel savings and tailpipe emissions of carbon dioxide (CO_2) (*1; 2*). In the first quarter of 2021, HEVs represented 18% of total passenger car sales in the EU, almost doubling their market share in a year (*3*).

The main purpose of using an HEV propulsion system is to decrease fuel consumption by using an internal combustion engine (ICE) assisted by an electrical motor (EM) to provide the required overall vehicle power output. The EM uses the energy stored in the batteries, which is either produced by the ICE or through regeneration from braking, to power auxiliary loads and reduce engine idling when the vehicle is stopped (*4*). HEVs are typically categorized according to their capability for full-electric driving (full HEV) or not (mild HEV) and divided in terms of powertrain configurations (parallel, series, power-split and multi-mode) (*5*). Optimization strategies for HEVs aim at minimizing fuel use or emissions, which in turn depend on the above HEVs types (*6; 7*).

Although many countries have been frequently using vehicle taxes and purchase subsidies to provide incentives to spur the electric vehicles (EVs) market, the penetration of EVs into the privately-owned passenger vehicle fleet is still lagging behind new vehicle sales. The forecasts for 2031 predict that EVs represent about 9% and 20% of all new vehicle sales in the USA and Europe, respectively (*8*), meaning that ICE and HEVs will still be the dominant in the worldwide vehicle fleet in the medium-term.

A good body of research has been devoted to exploring passenger car HEVs and conventional gasoline/diesel vehicles fuel use and emissions under real driving emissions (RDE) conditions. Two general observations hold across the state-of-the-art: 1) HEV fuel use and CO_2 are usually lower than a comparable gasoline ICE powered vehicle (*2; 9; 10*); and 2) Restart of the ICE results in sporadic spikes in gas-phase and ultrafine particle emissions in HEVs exceeding those obtained in hot-stabilized conditions (*10-15*).

Emission performance of HEV powertrains also depends on vehicle operating mode, ambient temperature, road grade, road geometry, congestion levels, or type of traffic control (*14; 16-20*). One of the first studies was conducted by Zhai et al. (*16*) who developed a modal emissions model for a 2001 Toyota Prius based on Vehicle Specific Power (VSP) and associated with startup and shutdown of the ICE. Alvarez and Weilenmann (*17)* examined the impact of low ambient temperature on CO_2 emissions for five HEVs models and revealed that the HEVs could reduce the cold-start extra emissions by 30% to 85% in comparison to a gasoline ICE vehicle. Ambient temperature is a relevant factor in the performance of a hybrid system battery, which in turn affects CO_2 emissions emitted by HEVs (*18*). A 2020 study operating one HEV passenger car on four roadway sections demonstrated the relevance of using real-world road grade as an input parameter in the estimation of CO_2 using a VSP model; the coefficient of determination (R^2) was 0.6 and 0.87 by using grade VSP and non-grade VSP models, respectively (*19*). More recently, a study on driving behavior (*14*) has reported significant variations in HEV-related fuel use and emissions under speed variations.

HEVs and EVs drivers very often adopt calm driving behaviors to obtain significant fuel economies (*21*), i.e., they tend to have less volatile driving styles than those of ICE vehicles. Volatility driving is due to reckless and aggressive driving styles that are associated with higher variations in acceleration/deceleration or higher

acceleration values (*10*), thus resulting in higher power requests to the vehicle powertrain. It can also depend on road type and traffic conditions related to traffic congestion or the presence of traffic incidents (*10*). Such driving styles have different impacts on fuel consumption and emissions even in identical test conditions (*22-24*).

Besides speed and acceleration/deceleration, vehicular jerk (the first derivative of acceleration) is often used to represent volatility in instantaneous driving decisions (*25*). Although transportation literature is rich in studying driving volatility on emissions under RDE *(14; 26)*, to the authors' knowledge, the relationship between HEV emissions, VSP, ICE operation status, and driving volatility according to the road type has not been explored before. A previous work conducted by the research team developed emission models for one 2019 SUV Toyota and associated driving behaviors with nine vehicular jerk types. Engine speed-based models showed to be good predictors of CO_2 and Particulate Matter (PM) by exhibiting R^2 higher than 0.70 for both ICE on/off states. However, the correlation of road sections with volatile driving behaviors and impacts on HEV emissions were not addressed (*14*).

Given the variability of HEV operation, a complete characterization of emission patterns and driving profile variability of speed and acceleration at road section level remains particularly challenging. Moreover, understanding the roadway characteristics and ICE operation status that cause certain volatile driving behaviors is of interest to further decrease fuel use and emissions in hybrid powertrains.

This paper fills these gaps by examining the unique data collected on tailpipe emissions, and vehicle engine and dynamics from a HEV sub-compact under RDE driving conditions, more specifically, second-by-second CO_2 and nitrogen oxides (NOx) emissions and six vehicular jerk types (*25*) on three road types (urban, rural and highway).

A modified principal component analysis technique with emissions, engine activity data, road characteristics, and vehicular jerk types as variables was applied in this study to extract principal components and identify specific clusters. In particular, these variables were evaluated using the Two-Step-Semidefinite programming (SDP) algorithm (*27*) based on the Clustering and Disjoint Principal Component Analysis (CDPCA) technique (*28; 29*). The CDPCA allows clustering of objects and simultaneously, partitioning of variables in a disjoint way. This represents an asset for interpretation of the relationships within the set of variables since each original variable contributes to a single principal component, and the set of objects, where each data point belongs to a single cluster (*29; 30*). The advantage of using the Two-Step-SDP algorithm relies on its remarkable performance when compared to the Alternating Least-Squares (ALS) algorithm: it revealed a significant improvement in terms of computational time, the proportion of variance explained by the disjoint components, and better recovering of both the true number of object clusters and the true variable partition (*30*).

2. RESEARCH QUESTIONS

This study seeks to answer the following key research questions:

1) How do CO_2 and NOx hot-stabilized emissions vary in RDE driving conditions according to the HEV ICE operation?
2) What are the differences between HEV-specific emissions and driving volatility hotspots on a specific RDE compliant route?
3) How does the level of volatility among vehicular jerk types change across three roadway types (urban, rural, and highway) and HEV ICE operation?

4) Which variables related to the emissions, dynamic and engine allow identifying hidden clusters in vehicular jerk types according to the route type and ICE status?

3. METHODOLOGY

3.1. Study Setup

The study collected tailpipe emissions, engine activity, and vehicle dynamics data from one 2020 Toyota Yaris HEV sub-compact with an initial odometer mileage of 2,500 km. The choice of the Yaris relied on the predominance of the well-known Hybrid Synergy Drive, which was used in 41% of HEVs sold in Europe in 2020 (*31*). The main characteristics of the test vehicle are listed in **TABLE 1**.

TABLE 1 Technical Specifications for the HEV used in this study

Specification	2020 Toyota Yaris
Engine	1.5L Three Cylinder
Fuel	Gasoline RON 95
Mileage (km)	2,500
Curb Weight (kg)	1,125
Actual Weight (kg)[1]	1,325
Transmission	Continuously variable transmission
ICE compression ratio	14:1
Combined power of EM plus ICE (kW)	85
ICE Horsepower (kW @ RPM)	67 @ 5,500
ICE Maximum Torque (Nm @ RPM)	120 @ 3,600
Electric motor (kW)	59
Electric motor torque (Nm)	141
Battery type	Lithium-ion
Number of electric engines	1
Maximum Speed ($km.h^{-1}$)	175
Emission Standard	Euro 6d
Combined Fuel Consumption (L/100km)	4.3
Combined WLTP CO_2 ($g.km^{-1}$)[2]	98

Note: 1) Sum of the curb weight of the vehicle plus the mass of instrumentation and personnel (200 kg); 2) WLTP is the Worldwide Harmonized Light-Duty Vehicles Test Procedure.

Tests were completed by one driver traversing a specified 44 km driving route over 4 days in April and May 2021 in Aveiro region (Portugal). The route was chosen because of its variability concerning the speed limits, traffic volumes, traffic control treatments (such as, roundabouts, traffic lights, stop-controlled intersections and priority intersections) and road grade, which is representative enough of the RDE driving conditions for a light-duty vehicle (*14*). The route typology was collapsed into three sectors (urban, rural and highway) based on speed limits and minimum distance, i.e., 13 km, as specified in **TABLE 1**.

FIGURE 1 depicts the selected route for this study, where point A represents the start of the trip. The rural sector corresponds to the beginning of the route (A-B) where the vehicle was driven with a speed lower than 90 $km.h^{-1}$. During the urban sector, the vehicle is driven at speeds lower than 50 $km.h^{-1}$ (B-C) in 93% of the trip

time. Concerning the highway, it includes the segment C-D, and with a speed higher than 90 km.h^{-1} in more than 60% of trip time. The selected route also includes a short urban sector (~1.7 km) corresponding to the segment D-A to conclude the trip. It is worthwhile noticing that traffic conditions and control treatments did not allow to perform all the sectors at the posted speed limits.

TABLE 2 Road Sector Summary

Road Sector	Distance (km)	Grade (%)		Number of Intersections			Maximum Speed Limit (km.h^{-1})	Time at speed > 50 km.h^{-1} (%)	Time at speed > 90 km.h^{-1} (%)
		min.	max.	R	TL	O			
Urban	13.1	-8.5	9.3	9	4	14	50	7	-
Rural	14.6	-12.4	12.3	9	2	-	90	65	-
Highway	16.2	-5.9	5.2	-	-	1	120	93	62

Note: min. = minimum; max. = maximum; R = Roundabouts; TL = Traffic Lights; O = Others.

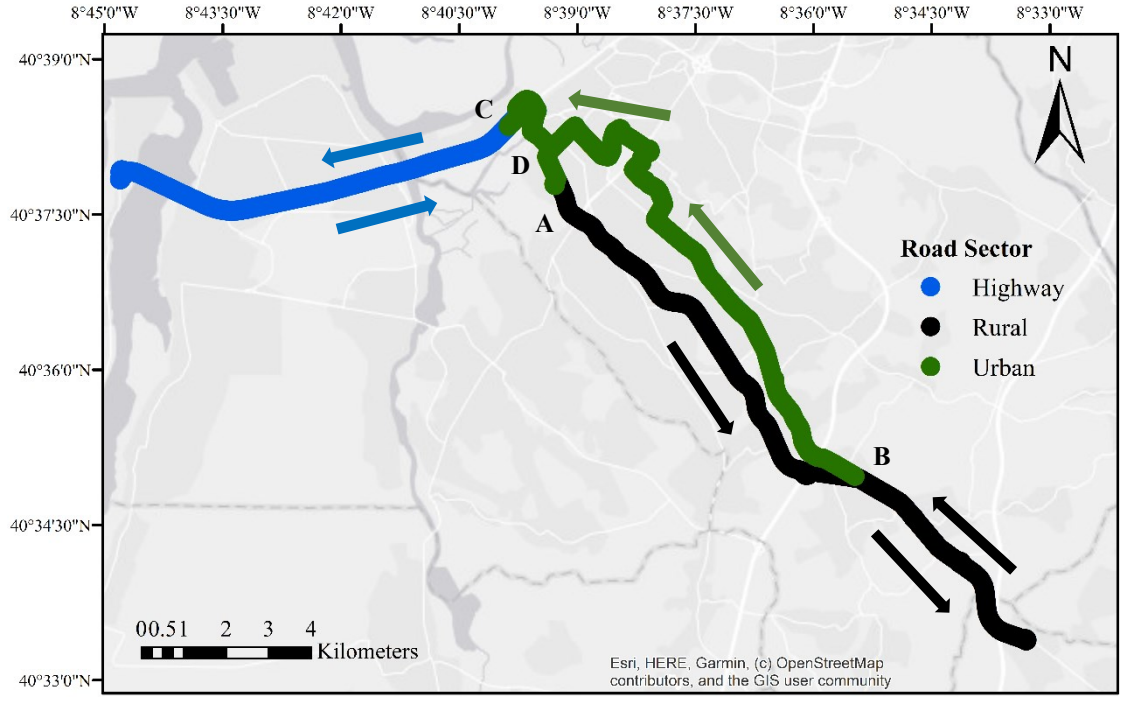


FIGURE 1 Specific route of RDE tests in Aveiro region (Portugal)

3.2. Tailpipe Emissions and Vehicle Data Collection

This paper used the 3DATX ParSYNC micro Portable Emission Measurement System (PEMS) (*32*) to collect CO_2 (in volume fraction with a range of 0-20%) and NO/NO_2 (with a range of 0-5,000 ppm) concentrations of exhaust pulled directly from the tailpipe, sampling at a rate of 1 Hz. NOx concentration was defined as the sum of concentration signals of nitric oxides (NO) and nitrogen dioxides (NO_2) (*33*). Although micro-PEMS are not used for regulatory certification testing, they are capable of measuring HEV emissions (*14; 34*) and identifying different driving behavior profiles with reasonable accuracy (*35*).

Trip tests preparation included routine calibrations of gas analyzers (zero and span drift) through the use of the UN 1956 gas mixture. During the ParSYNC

experimental routine, a warm-up phase of around 15 km preceded all trips, allowing for the engine coolant temperature to stabilize on values greater than 80ºC (*20*).

Neither parSYNC has an exhaust flowmeter nor includes an internal On-board Diagnostic (OBD) reader. Thus, a Bluetooth OBD-II was connected to the electronic car unit (ECU) to record at a frequency of 1 Hz the following engine parameters:

- OBD speed ($km.h^{-1}$);
- Engine Speed (RPM);
- Mass Air Flow (MAF) ($g.s^{-1}$);
- Fuel Flow Rate (FFR) ($L.s^{-1}$);
- Engine coolant temperature (ºC).

A QSTARZ GPS Travel Recorder (position accuracy of 3 m) to get second-by-second vehicle position and elevation.

A single male driver (27 years old) was used in all sampling runs [a total of 6, according to the methodology described here (*34*)] as an intentional study design with the main purpose of creating a similar driving style across the road sectors. The weather conditions observed during driving sessions were characterized by ambient temperatures between 19 °C and 22 °C, and relative humidity lower than 70%.

3.3. Data processing and Quality Assurance

Time alignment of PEMS, OBD and GPS data was done before computation of mass emissions following the methodology described by Sandhu and Frey (*33)*. The alignment method was based on pair of variables exhibiting concurrent trends (e.g., rise in MAF and concordant rise in the concentration of exhaust gas) and further evaluated using the Pearson Correlation Coefficient (PCC), which measures the degree of linear correlation between two variables. The use of engine RPM and NOx concentration as pairs to be synchronized in gasoline powertrains is recommended (*33*). However, HEV collected data do not allow to visualize a concurrent trend between these variables due to the continuous switch between ICE and EV modes and negative concentrations of NO and NO_2. Therefore, the CO_2 volumetric concentration and FFR were used. The process was repeated every day the PEMS was started. The maximum PCC values related to the CO_2-FFR synchronized pairs were higher than 0.84.

Frequently used indicator variables in synchronizing OBD and GPS data are OBD speed versus vehicle speed. These variables produced PCC peak values of approximately 0.997, which means that data sets were properly synchronized (*33*). GPS data allowed to obtain elevation profile, and then resulting road grade (*r*) in each 100 m-length was computed following the segment method suggested in (*36*).

Erroneous data were checked by taking into account quality assurance screening criteria (*33*). Common errors found during the monitoring campaigns were related to the negative NO and NO_2 concentrations, which represent values below the instrument detection limit. Such negative concentration values were zeroed for emission analysis.

3.4. Computed Parameters

3.4.1. Vehicle Specific Power

VSP represents the vehicle instantaneous power per unit mass, and it accounts for engine power demand associated with changes in both vehicle potential and kinetic energies, aerodynamic drag, and rolling resistance (*37*). The second-by-second collected data of OBD speed (v_i) and acceleration (a_i), and road grade (*r*) were used to quantify

instantaneous VSP with assumptions for light-duty vehicles constants based on the work of Jimenez (*37*), as denoted by Eq. 1:

$$VSP_i = v_i\left[1.1a_i + 9.81r + 0.132\right] + 0.000302.v_i^3, \quad (1)$$

where:

VSP_i = Vehicle Specific Power in the second of travel i (kW.ton^{-1});

3.4.2. Mass Emission Rates

Second-by-second CO_2 and NOx emissions rates (mass per time) were computed based on the Regulatory Information 40 CFR 86.144 for tailpipe emissions (*38*). The exhaust mass flow rate from data as reported by the ECU (*14*) can be computed as follows (Eq. 2):

$$\dot{m}_{ex} = MAF + FFR.\rho_{fuel}, \quad (2)$$

where:

$\dot{m}_{ex}$ = Exhaust mass flow rate (g.s^{-1});
MAF = Mass air flow rate measured by the OBD (g.s^{-1});
FFR = Fuel flow rate measured by the OBD (L.s^{-1});
ρ_{fuel} = Fuel density (730 kg.L^{-1} at 15°C).

CO_2 and NOx mass emission rates (corrected to standard conditions) are then obtained through the expression in Equations 3 and 4, respectively(*38*):

$$m_{CO_2} = \dot{V}_{exhaust}.\rho_{CO_2}.X_{CO_2}, \quad (3)$$

$$m_{NOx} = \dot{V}_{exhaust}.\rho_{NOx}.X_{NOx}.\frac{1}{1-0.0047\left(H-75\right)}, \quad (4)$$

where:

$\dot{V}_{exhaust}$ = Exhaust volumetric flow rate (corrected to standard conditions) (m^3.s^{-1});
ρ_{CO2} = Density of CO_2 at the standard conditions (1.830 kg.m^{-3});
X_{CO2} = Volume fraction of CO_2 measured by PEMS (%);
ρ_{NOx} = Density of NOx at the standard conditions (1.913 kg.m^{-3});
X_{NOx} = Volume fraction of NOx measured by PEMS (ppm);
H = Humidity (%).

3.4.3. Vehicular Jerk Types

Instantaneous driving decisions imply short-term driving decisions to accommodate real-time changes during the trips, including pavement conditions, approaching traffic control treatments, or the presence of adjacent vehicles, which are common under RDE conditions. Wang et al. (*25)* reported that instantaneous driving decisions are related to acceleration/deceleration, constant speed (zero acceleration), jerking the vehicle (rate of change acceleration or deceleration) and/or maintaining constant acceleration and deceleration (zero vehicular jerk) episodes. Vehicular jerk (*j*) is the derivative of acceleration with respect to time, being considered as a proper kinematic variable for capturing drivers' abrupt adjustments (*25*).

Six different types of vehicular jerk profiles can be defined based on acceleration and deceleration profiles at each sampling speed. Such instantaneous driving behaviors include jerk enhancements (types B and D), jerk reversals (types C and F), and jerk mitigations (types A and E), as follows (*25*):

A) Acceleration followed by lower acceleration: $a_i >= 0$ & $a_{i+1} > 0$ & $a_i > a_{i+1} \rightarrow j < 0$;
B) Acceleration followed by higher acceleration: $a_i >= 0$ & $a_{i+1} > a_i \rightarrow j > 0$;
C) Acceleration followed by deceleration: $a_i >= 0$ & $a_{i+1} < 0 \rightarrow j < 0$;
D) Deceleration followed by lower deceleration: $a_i < 0$ & $a_i < a_{i+1} < 0 \rightarrow j > 0$;
E) Deceleration followed by higher deceleration: $a_i < 0$ & $a_{i+1} < a_i < 0 \rightarrow j < 0$;
F) Deceleration followed by acceleration: $a_i < 0$ & $a_{i+1} > 0 \rightarrow j > 0$

Noted that a_i and a_{i+1} represent acceleration ($m.s^{-2}$) in the second of travel i and i+1, respectively.

3.5. Two-Step-SDP algorithm

To detect the most relevant information behind emission and driving volatility data, the CDPCA methodology (*28*) is of special interest because it allows partitioning the variables into a reduced set of disjoint components (i.e., each component is characterized by a disjoint set of variables) and unveiling patterns among the objects through clustering objects along with a set of centroids, simultaneously, so that the between cluster deviance of the components in the reduced space is maximized. The CDPCA model results from applying the K-means on the data matrix for clustering the objects and simultaneously, performing sparse Principal Component Analysis (PCA) on variables, which leads to an improvement of interpretability (*28; 30*). The CDPCA model can be given by Eq. 5:

$$X = U\bar{Y}A^T + E\,, \tag{5}$$

where:

X = Data matrix ($I \times J$);
U = Object membership assignment matrix ($I \times P$);
A = Orthonormal components loadings matrix ($J \times Q$); $\bar{Y} := \bar{X}A$, where $\bar{X}$ is the ($P \times J$) object centroid matrix in the original space, represents the ($P \times Q$) cluster centroid score matrix;
E = Error matrix associated with the model ($I \times J$).

An iterative heuristic procedure suggested in (*29*) and an approximation algorithmic framework based on SDP proposed by (*27*) were proposed to solve the CDPCA problem. The latter is called the Two-Step-SDP algorithm and is based on SDP relaxations of two clustering problems and a K-means step in the reduced space (*27*). The Two-Step-SDP approach outperforms the ALS yielding improvements not only regarding the computational time but also in terms of the proportion of variance explained by the disjoint components (*30*). This algorithm also reveals better recovering of both the true number of object clusters and the true variable partition clustering assignments (*30*).

The Two-Step-SDP algorithm for CDPCA involves a first phase where the P clusters of objects and Q clusters of attributes are initially estimated considering

orthogonal projections and 0-1 SDP models. These SDP models are relaxed to convex models and solved using a Singular Value Decomposition-based approach. After that, a rounding procedure based on K-means applied in the reduced space of centroids is performed to obtain the CDPCA model solutions. More details about the procedure can be found in (*27*).

4. Results

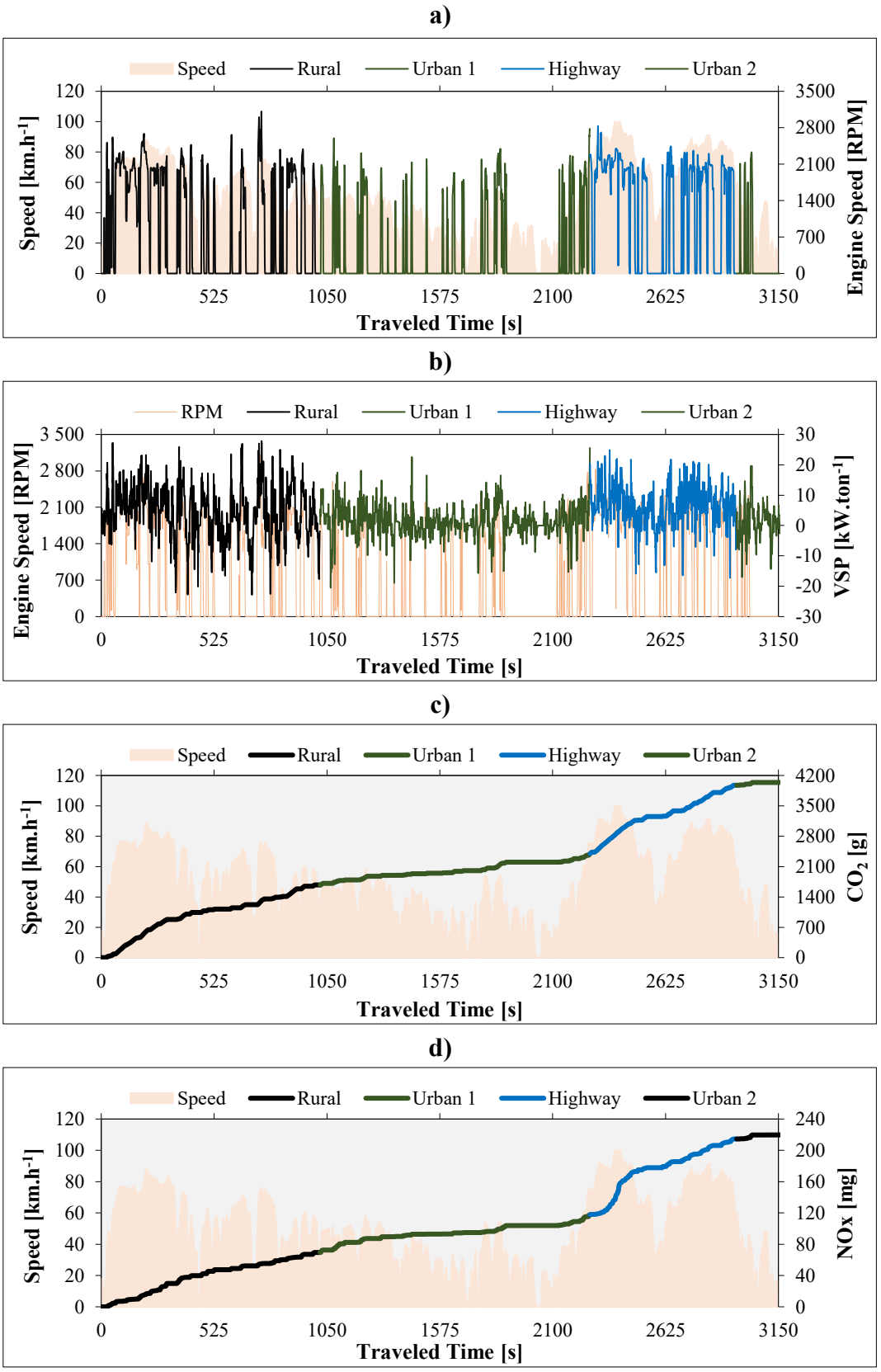
4.1. Analysis of engine performance and emissions

Collected data covered approximately 15,500 s and almost road coverage of 265 km. Validation of the trips confirmed that all trips conducted in urban, rural and highway road sectors fulfilled the RDE criteria related to the relative positive acceleration and the ninety-five percentile of the product of 1 Hz of vehicle speed and positive accelerations higher than 0.1m.s^{-2} (*14*).

From these results, values of average mass trip emissions and standard deviation were computed. CO_2 emissions per kilometer (100 ± 3 g.km^{-1}) are close to the vehicle-type approval, as reported in **TABLE 1**. Further break down of the CO_2 performance across three road sectors is more evident under urban (63 ± 12 g.km^{-1}) than rural (105 ± 5 g.km^{-1}) and highway (126 ± 7 g.km^{-1}).

FIGURE 2 a-d shows a HEV trip driving sample in terms of instantaneous OBD Speed, RPM, ICE mode and VSP, and cumulative CO_2 and NOx. The first part of the trip corresponds to the road sectors in rural and urban roads, and, under these driving conditions, the engine speeds are significantly different because of the cyclical on-off states of the ICE. This is especially true in the urban sector where resulting values of cumulative CO_2 increase slowly throughout its length. In the HEV highway driving, noticeable increments for both CO_2 and NOx emissions are obtained, which are expected to be a consequence of the acceleration episodes during the highway entrance ramp and high-vehicle power demand in uphill road sections. For all road sectors, the expected relationship between ICE on-off and VSP is observed; RPM is typically 0 for VSP values equal or lower than 0 kW.ton^{-1} (idling, braking, and low-speed operation). Scatter plots also revealed that the electric mode of the HEV only occurs up to VSP values of 15 kW.ton^{-1}, which is coherent based on available EM power alone and the testing vehicle weight (**TABLE 1**).

The distribution of FFR, CO_2, NOx, distance, and travel time by road sector is exhibited in **FIGURE 3**. The values in the graphs represent the average values of all trips performed by HEV. The highway road sector contributes to the greatest portion of the CO_2 and NOx emissions (respectively, 42% and 54%) generated by a vehicle along the route. This segment corresponds to about 24% of the travel time. The expected fuel-use and emissions reductions attained from driving HEV technology are confirmed in urban road sectors, as already observed in previous studies (*2; 9; 10; 14*). They account together with less than 20% of the fuel use and emissions in more than 50% of trip time. Overall, the proportion of time in electric-drive-only operation (EV) was higher during the urban sectors (76%), followed by rural (50%) and highway (28%) sectors.



Note: Urban 1 corresponds to the segment B-C; Urban 2 corresponds to the segment D-A

FIGURE 2 Example of driving route from a single trip (Trip 1) of the HEV: a) OBD speed and RPM; b) ICE mode (On = 1; Off = 0) and VSP; c) OBD speed and cumulative CO_2; and d) OBD speed and cumulative NOx

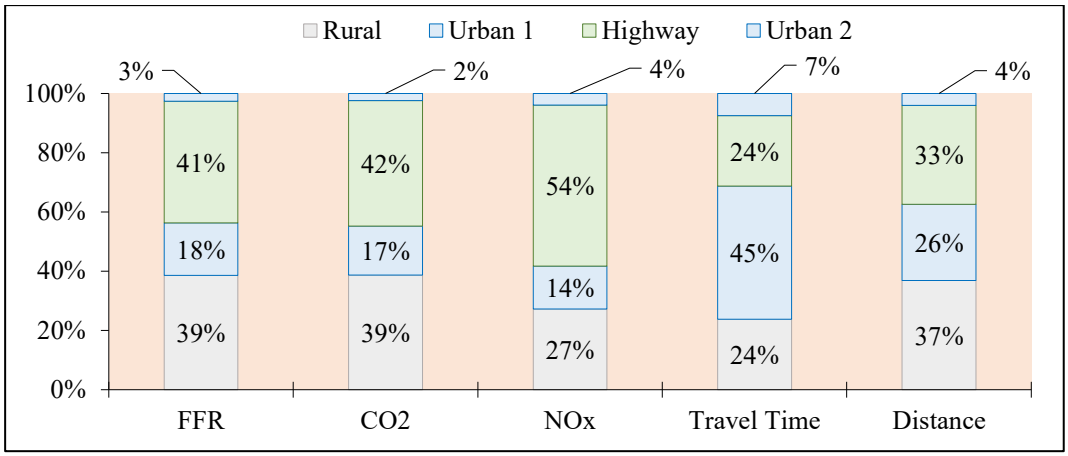


FIGURE 3 Distribution of FFR, CO_2, NOx, travel time and distance of the HEV by road sector

To identify the locations with higher emissions in each road sector, a distribution of CO_2 and NOx emissions during the route is depicted in **FIGURE 4** a-b. Each point represented in the map indicates the average, computed as the total mass of emissions generated in each 100-m length divided by the number of trips. Additionally, the points in green are the segments where ICE was off. The highest values are clearly distinguished during rural and highway sectors exhibiting, for instance, CO_2 values higher than 11 g in more than 45% of their segments and NOx values higher than 1.89 mg in more than 10% of their segments. A close view of the urban road sectors reveals several locations with no CO_2 emitted by the HEV (EV mode), which accounted for 20% of segments in both road sectors. It was also possible to identify some locations in EV mode on the rural and highway sections, which can be justified by two main reasons: regenerative braking at low speed and/or deceleration and the existence of downgrade segments.

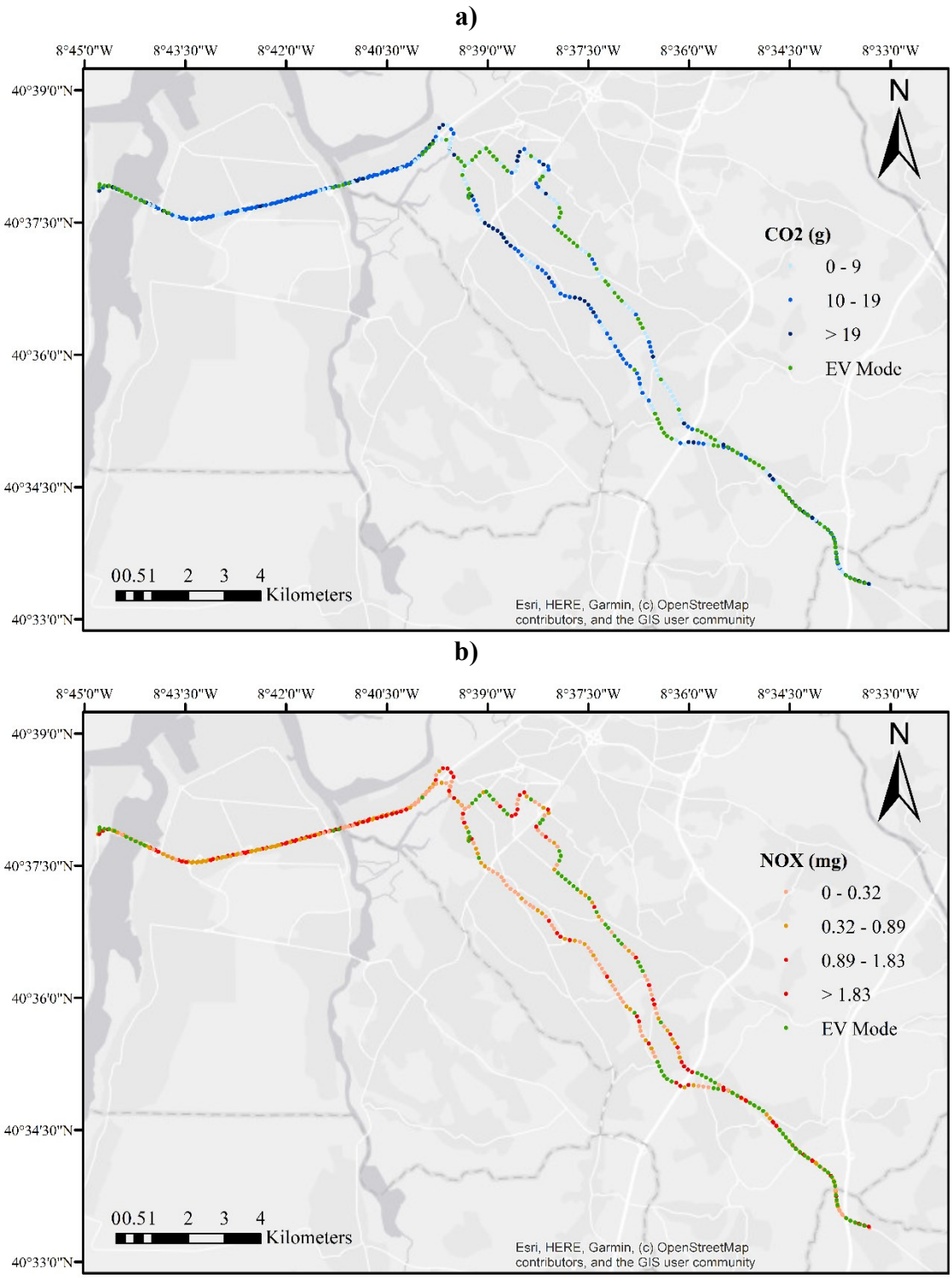


FIGURE 4 Emissions in 100m-interval along route: a) CO_2; and NOx

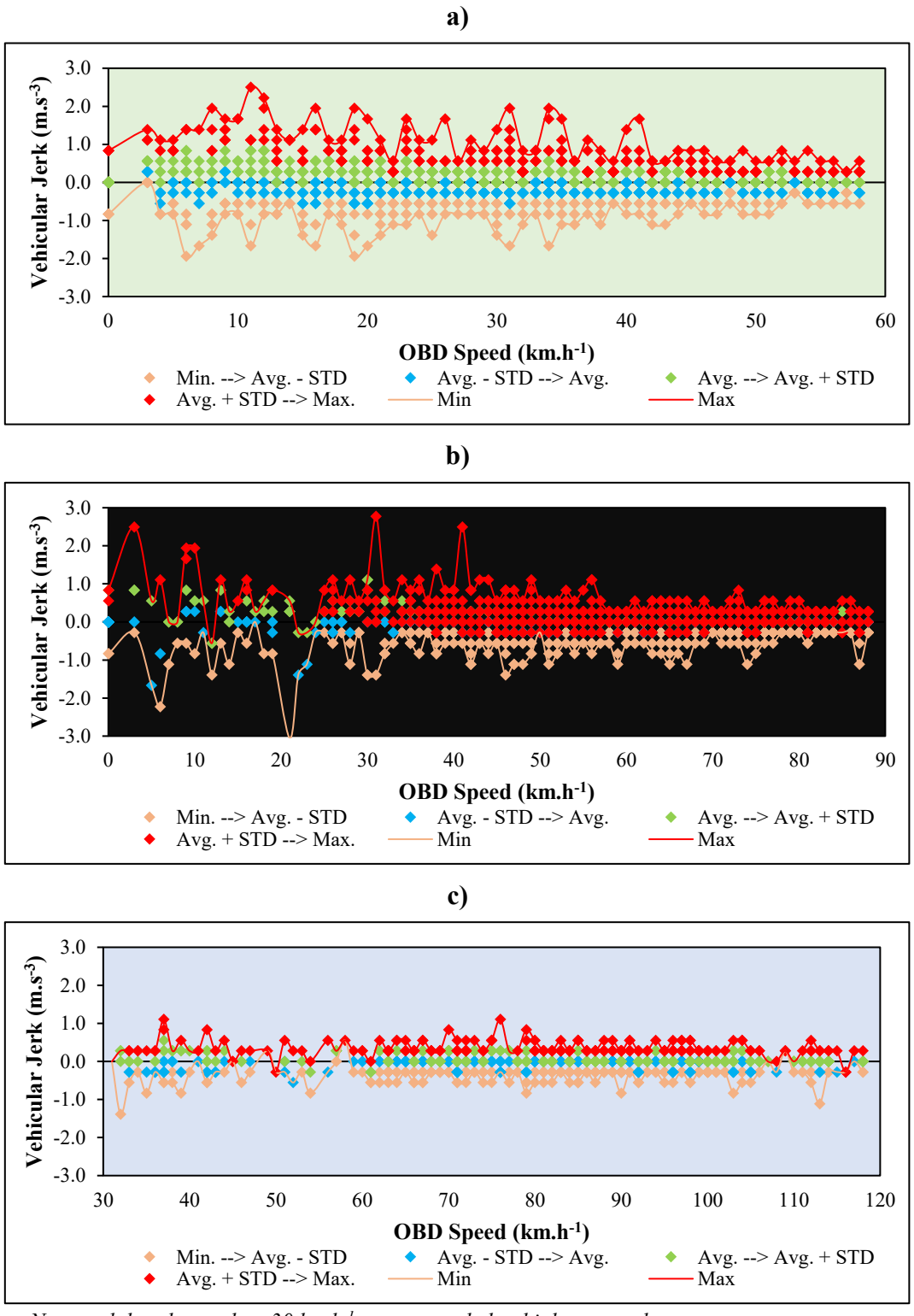
4.2. Analysis of driving volatility indicators

To examine the typical profile of regular driving volatility by road type, vehicular jerk by speed bins in 1 $km.h^{-1}$ increments is calculated based on second-by-second data set (**FIGURE 5** a-c). Scatter plots also include upper and lower bands (mean plus or minus standard deviation) for the aggregated positive and negative vehicular jerk values. Upper band characterizes the most typical driving practice on the roadway. The red and

orange points that are out of the bands are defined as “highly” volatile driving behaviors while blue and green points represent “moderately” volatile driving behaviors (*25*).

Urban driving covers a wider band of vehicular jerks when compared to the other road sectors. About 62% of the data samples of urban driving is highly volatile (44% and 17% for a positive and negative vehicular jerk, respectively). Low-speed data (< 45 $km.h^{-1}$) exhibit larger bandwidth than high speeds data did. The rural road sector yields 28% of the data sample as highly volatile (15% and 13% for a positive and negative vehicular jerk, respectively), meaning that the typical driving behavior is less volatile under these conditions. For speed values higher than 50 $km.h^{-1}$, the absolute vehicular jerk values are generally lower than 1 $m.s^{-3}$. The largest bandwidth is between 10 to 40 $km.h^{-1}$ under highway driving and it decreases markedly when speed is higher than 60 $km.h^{-1}$. The latter speed intervals are represented by vehicular jerk values lower than 0.3 $m.s^{-3}$, as reflected in the upper band.

To complement the analysis, a mapping of the vehicular jerk data for two samples is depicted in **FIGURE 6** a-b. For clarity of comparison, the dataset is divided in terms of negative, zero and positive jerking. The number of seconds in vehicular jerking (orange and red points) is notably high in urban driving conditions, which is reasonable because vehicles are continuously braking and accelerating due to the presence of intersections, other vehicles, and traffic incidents on the road. Rural and highway show more data corresponding to constant acceleration and speed (zero jerking), especially in sections characterized by fewer stop-and-go traffic where vehicles are allowed to move in free-flow conditions and with smoother variations in acceleration. These findings are relevant when combined with information given in **FIGURE 5**. They confirm that the trend observed in vehicular jerk profile widely varies according to the type of road, which in turn can explain specific values on tailpipe emissions and engine variables of HEV.



Note: No speed data lower than 30 km.h^{-1} were recorded at highway road sector

FIGURE 5 Vehicular jerk distribution by speed bins: a) urban; b) rural; and c) highway

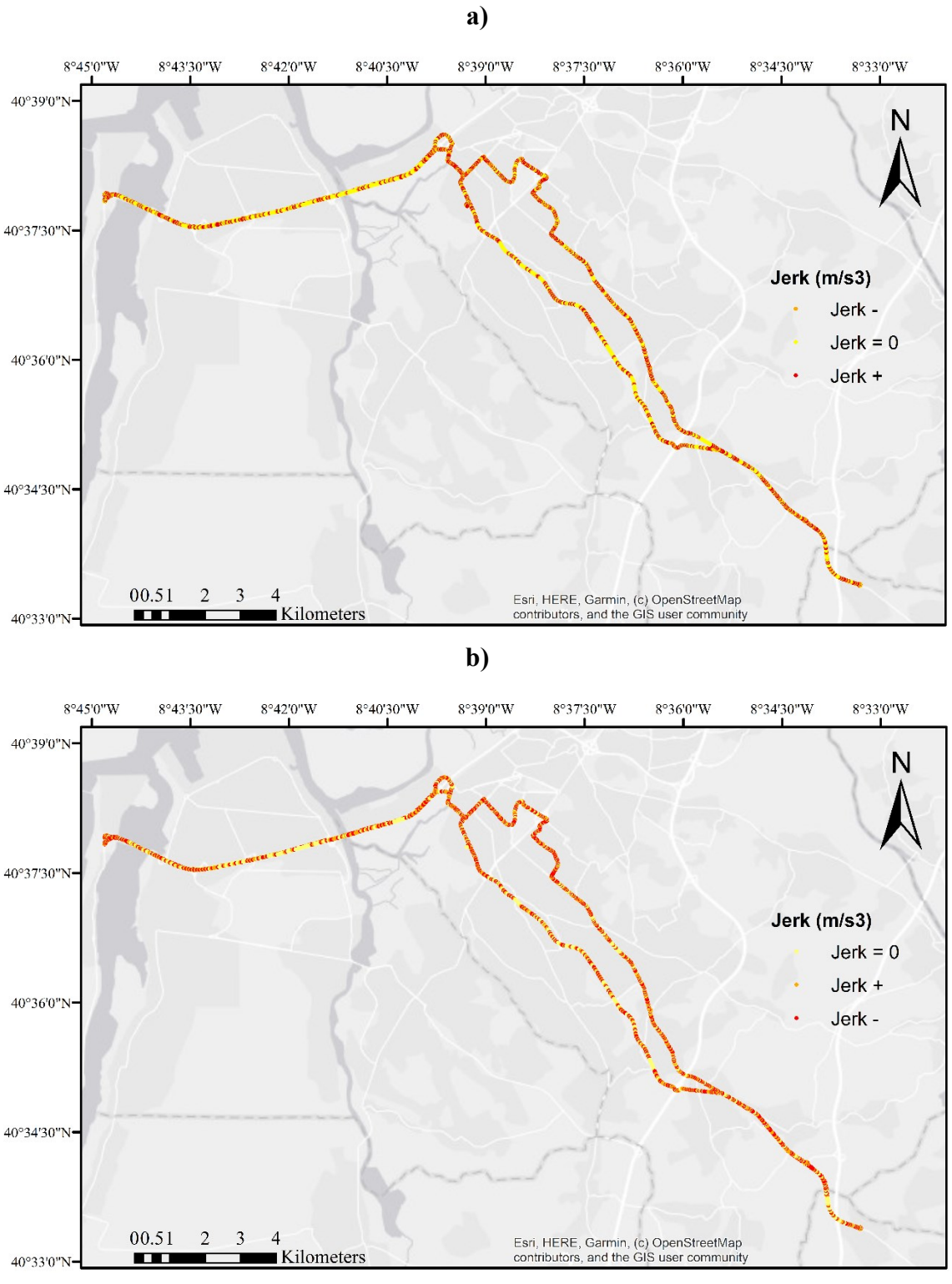


FIGURE 6 Identification of driving volatility behaviors by road sector: a) Trip 1; and b) Trip 4

To understand how much time is spent on different driving volatility states, the time spent for speed bins is aggregated by vehicular jerk type and road sector, as shown in **FIGURE 7** a-c. Because the number of data points is small (less than 10% of the dataset), analysis is centered in speed values higher than 40 $km.h^{-1}$ and 60 $km.h^{-1}$ in

rural and highway, respectively. The percentage of time spent on zero vehicular jerking yields the highest portion, especially in locations outside urban environments. These conditions of constant acceleration accounted for 19%, 29% and 41% of driving time at urban, rural and highway, respectively. When the vehicle is stopped, the zero jerking can represent more than 85% of the sample collected in the urban road sector. Jerk types A and B show significant amounts of time spent on urban and rural roads with approximately 20% each. Driver maintaining vehicle acceleration (D and E) represents a small portion of the dataset (between 4 to 12%, depending on the road sector and jerk type). However, urban driving is characterized by a percentage of these volatile states that were higher than those observed in other road environments.

FIGURE 8 a-b exhibits the distribution by jerk type and HEV ICE operating mode. Zero vehicular jerks represent more than 30% and 40% of the rural and highway data set, respectively, related to the ICE status "on". However, jerk types A and B are the most relevant ones under urban driving in the range of ~30-36%. Driver reversing vehicle acceleration (C and F) and maintaining deceleration (D and E) account together for less than 15% of the driving time, regardless of the road sector. The analysis of vehicular jerk for EV mode data dictates a different distribution among jerk types. The percentage of time spent on zero vehicular jerks remains to the highest portion of the rural and highway, but its relative contribution decreases to less than 30% of the driving time along with these road sectors.

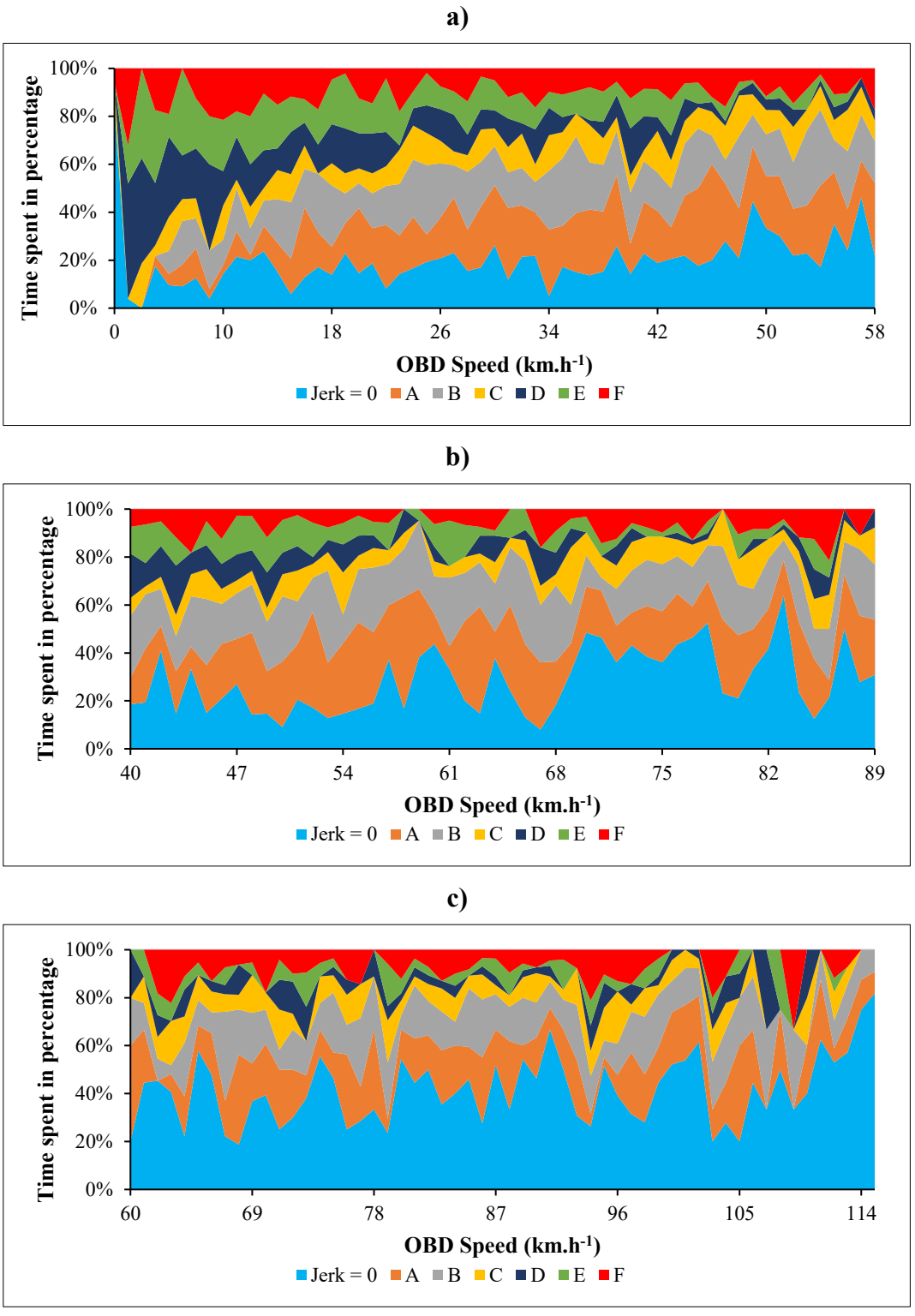


FIGURE 7 Time spent in different jerk styles: a) urban; b) rural; and c) highway

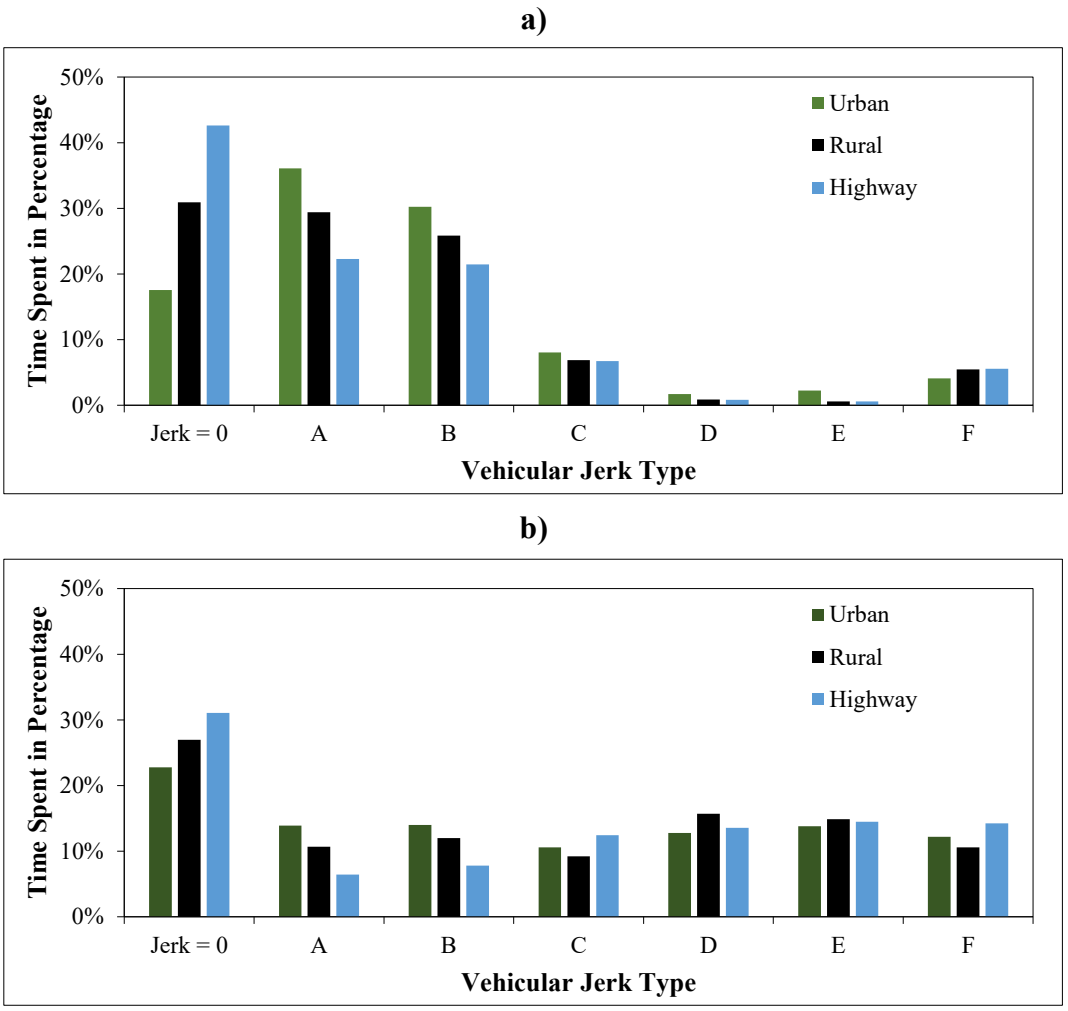


FIGURE 8 Time spent in different jerk styles by road sector: a) ICE mode; and b) EV mode

4.3. Numerical results of CDPCA

The Two-Step-SDP algorithm was applied to the above data set to perform CDPCA and identify relevant factors contributing to explaining the variability and reveal the cluster structure hidden on data. Two cases were considered: 1) data set constrained to the road sector and considering $P = 2$, which is related to the ICE and EV modes; 2) data set conditioned to the road sector together to a constraint concerning the operating mode (ICE or EV), and in this case, $P = 7$, which is related to the jerk type classes.

In all experiments, the tolerance was set to 10^{-5}, the maximum number of iterations was 100 and three principal components ($Q = 3$) were considered. CDPCA was performed for each road sector data points with RPM, OBD speed, acceleration, vehicular Jerk, VSP, vehicular Jerk Type, CO_2 and NOx chosen as variables.

4.3.1. Case 1

The algorithm converged to an optimal solution after 7, 8, and 8 iterations for the urban, rural and highway cases, respectively. The three components of the CDPCA explain almost 70% of the data variability regarding the urban and highway data, while for the rural case, this value is around 56%. RPM can be considered a variable that can explain the variability within the ICE or EV mode. For the urban case, the four factors

contributing to the first Component 1 are RPM, VSP and CO_2 and NOx emissions and can be identified to reflect the distinctive differences between ICE operation status. The RPM, acceleration and VSP seem to contribute most for the differences on the operating mode at rural sector, while for highway sector, the most relevant factors contributing to the first principal component are RPM, speed, vehicular jerk, and CO_2. **FIGURE 9** homogeneous clusters and a clear separation of the ICE and EV modes; the between cluster deviance of CDPCA is 55% for urban, and 63% for rural and highway sections of total deviance. Each road sector exhibits a different clustering structure. In the urban case, the cluster regarding the ICE mode presents a significantly thinner and longer shape when compared to EV mode. In the rural section, the clusters are well separated and more balanced in terms of shape. Finally, the highway sector presents clusters that allow identifying which points belong to the ICE or EV modes.

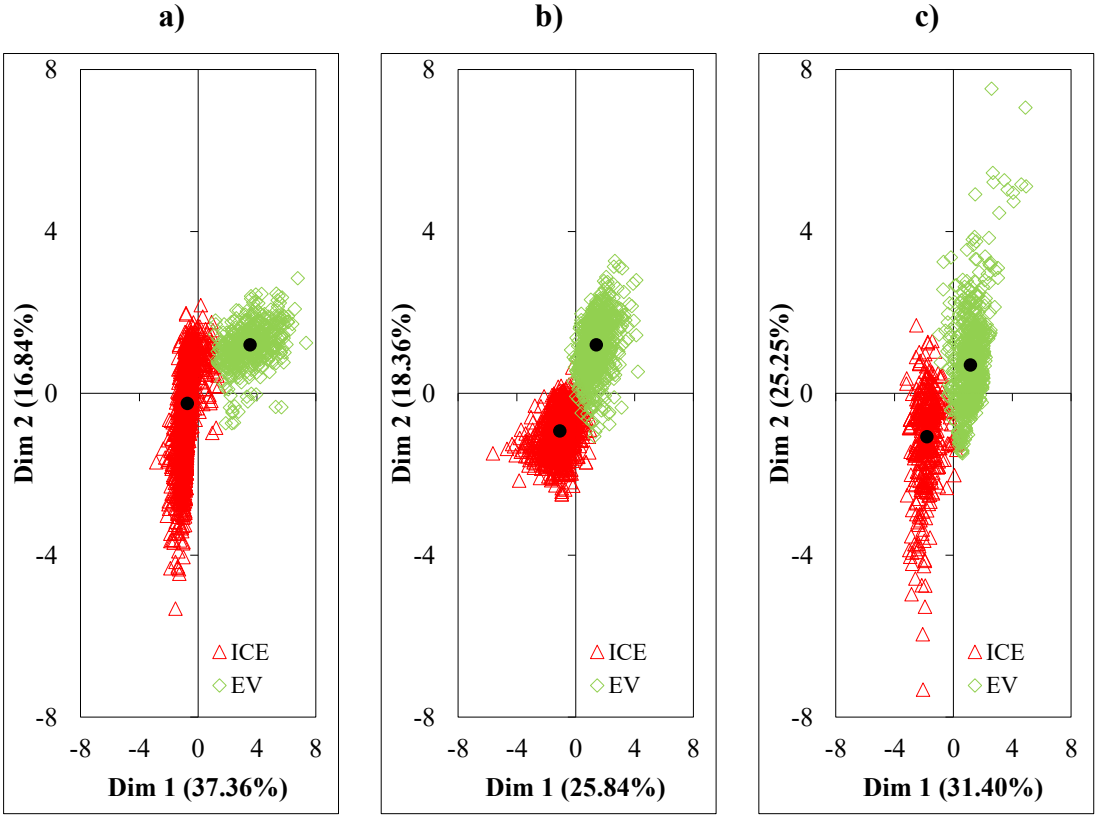


FIGURE 9 CDPCA results on the operating mode constrained to road sector: a) urban; b) rural; and c) highway

4.3.2. Case 2

The optimal solution for ICE mode was found after 28, 23 and 36 iterations for the urban, rural and highway cases, respectively. The three obtained components explain between 71% and 77% of the total variance for the urban, rural and highway cases. RPM seems to play a significant role in reflecting the differences between jerk types for all cases. The explained variance obtained by Component 1 referring to the RPM, VSP and CO_2 and NOx is higher than 41% and these four variables can be used to distinguish the jerk type for the urban sector. While for the urban case, the VSP contributes to the first principal component and other relevant acceleration-dependent factors contribute to the second component, for the rural and highway sections, these

present a similar composition of the most relevant factors that are related to acceleration-dependent variables and pollutant emissions. Under Component 1, RPM, Speed and CO_2 account for approximately 30% of the total explained variance for rural, that together with the second component reaches more than 55%, where the contributors are those acceleration-dependent variables (acceleration, jerk and VSP). The Two-Step-SDP algorithm could return a clustering of the objects with a clear separation by the jerk types (**FIGURE 10**); the between cluster deviance of CDPCA is clearly above 70% of total deviance for all road sections.

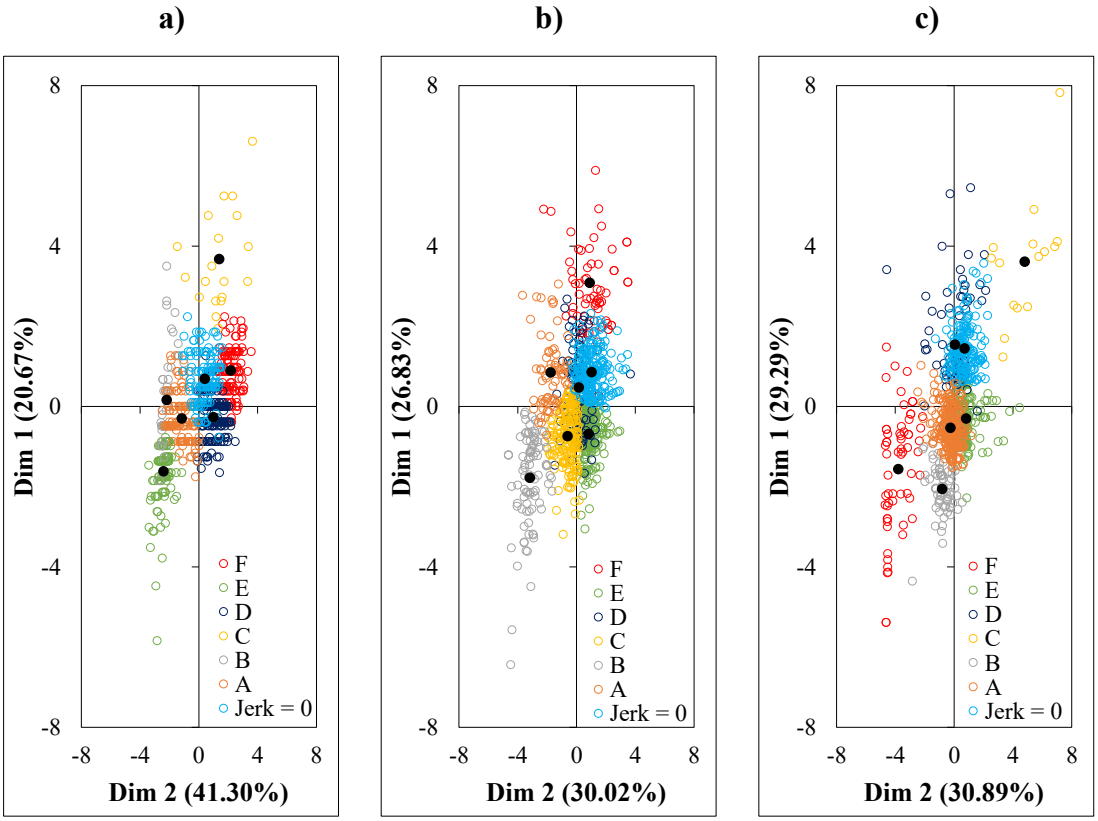


FIGURE 10 CDPCA results on the jerk types constrained to road sector and ICE mode: a) urban; b) rural; and c) highway

The optimal solution for EV mode was found after 28, 52 and 25 iterations for the urban, rural and highway sectors, respectively. The first two components explain more than 58% of the total variance for the urban, rural and highway data. Results reveal that the variables that contribute to each one of the three principal components coincide in all road sectors. The explained variance obtained by Component 1 referring to RPM and CO_2 and NOx is between 31-34% (depending on the road sector). Component 2 includes in all cases, the acceleration-dependent variables and the explained variance is approximately 28-30%. For all cases, the between cluster deviance of CDPCA is around 80% of total deviance. **FIGURE 11** highlights the clustering results after performing the CDPCA to try to identify the patterns by jerk types. When comparing the ICE and EV results (**FIGURE 10**), a clear difference in the shape of the clusters with respect to the jerk types can be observed.

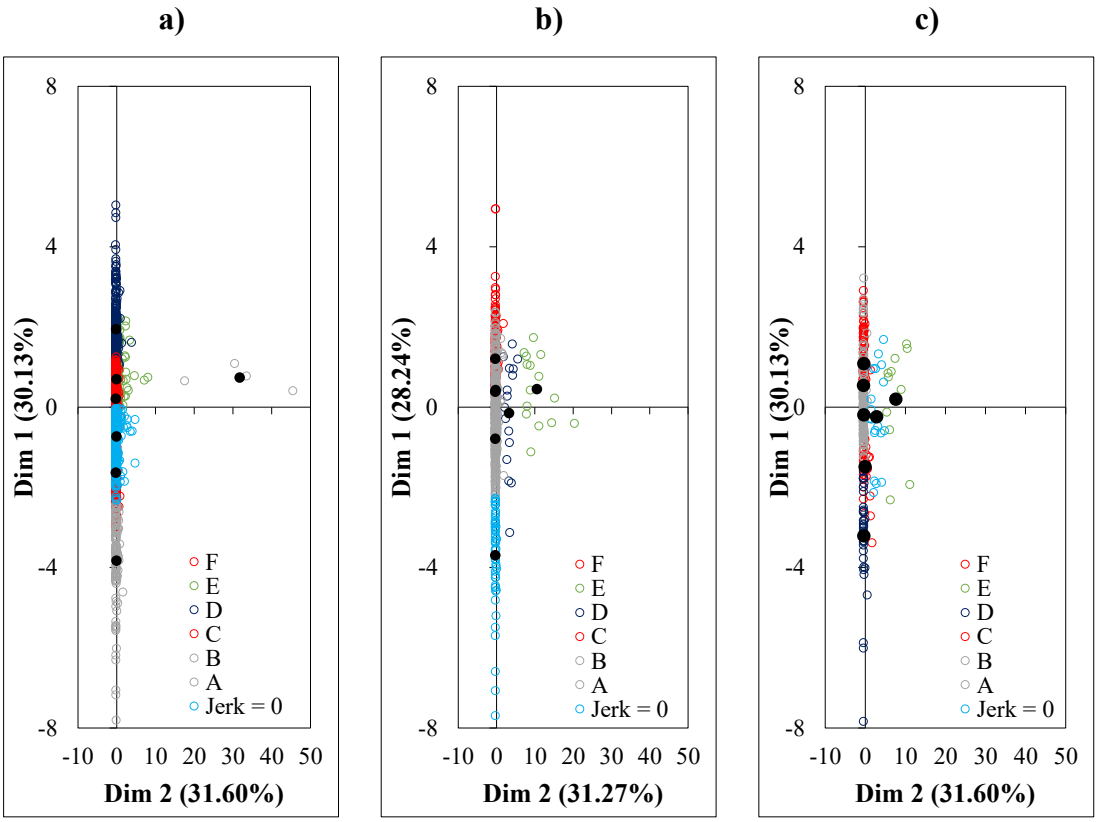


FIGURE 11 CDPCA results on the jerk types constrained to road sector and EV mode: a) urban; b) rural; and c) highway

5. CONCLUSIONS

This research focused on the comprehensive understanding of tailpipe emissions, vehicle engine and driving volatility of an HEV in different road sectors (urban, rural and highway) and for both ICE status. Driving volatility was categorized by several driving behaviors associated with a vehicular jerk that included zero jerking (constant speed and/or acceleration), jerk enhancements (increasing accelerations or decelerations), jerk reversals (acceleration followed by deceleration), and jerk mitigations (decreasing accelerations or decelerations).

Average CO_2 tailpipe emissions by road sector varied considerably from 20% of total urban route operation to 39% of total rural operation and 42% of total highway route operation. Characterization of ICE status found that average electric-drive-only operation was between 28% and 76% at highway and urban road sectors, respectively. Analysis of vehicular jerk showed the largest bandwidth of vehicular jerk under urban driving conditions, especially at speeds lower than 45 $km.h^{-1}$. The distribution of vehicular jerk widely differed across road sectors. Zero jerkings represented 19%, 29% and 41% of driving time at urban, rural and highway, respectively. Results also indicated discrepancies concerning the distribution of vehicular jerk types between ICE and EV modes. CDPCA results constrained to road sector confirmed that RPM, VSP, CO_2 and NOx are factors that allow identifying the cluster structure hidden on vehicular jerk type and ICE operation status data.

These findings highlighted the importance of measuring driving volatility when evaluating HEV in the context RDE analysis. The analysis of driving behavior based on the extent of driving volatility, which goes beyond simply labeling a driver as

aggressive or non-aggressive, and its further correlation with emissions are thus key study contributions. Vehicular jerk classification can be integrated into driving behavior monitoring and feedback devices able of using volatility information to provide alerts and warnings according to the ICE operation status. Applications can be embedded in vehicle navigation systems to warn drivers when high CO_2 and NOx emission and jerking movements are detected during a trip. This research has also scientific contributions in what respects which variables can be used as an indicator to distinguish and identify different clusters considering the jerk types between ICE operation status using Two-Step-SDP algorithm for performing CDPCA.

Although the methodology presented in this research can be used in any HEV or ICE, data used in this paper are from one HEV, so these conclusions are only applicable to similarly operated vehicles. Other HEVs may be operated following different optimization strategies and might reveal a different influence of road type characteristics and engine variables on vehicular jerk distributions and tailpipe emissions. Future work will be devoted to the characterization of HEV-specific CO_2, NOx and PM based on different driving behavior styles, engine operating temperatures (cold- and hot-stabilized exhaust emissions) and exhaust gas treatment systems. Next research steps will also focus on the development of a graphical interface capable of remotely connecting with ECU system to both incorporate and map vehicular jerk classification, as well as to define a driver index to measure the extent of variations and emission rates in driving.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the projects: UIDB/00481/2020 and UIDP/00481/2020 - Fundação para a Ciência e a Tecnologia; and CENTRO-01-0145-FEDER-022083 - Centro Portugal Regional Operational Programme (Centro2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund; DICA-VE (POCI-01-0145-FEDER-029463), inFLOWence (POCI-01-0145-FEDER-029679) and Driving2Driverless (POCI-01-0145-FEDER-031923) projects, co-funded by COMPETE2020, Portugal2020-Operational Program for Competitiveness and Internationalization (POCI) and European Union's ERDF (European Regional Development Fund); and "PAC Portugal AutoCluster for the Future" project, funded by PORTUGAL 2020 Partnership Agreement. R. Tomás would like to acknowledge the support of FCT for the Ph.D. Scholarship 2020.07968.BD.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: P. Fernandes, R. Tomás, E. Macedo; data collection: P. Fernandes, R. Tomás; analysis and interpretation of results: P. Fernandes, R. Tomás, E. Macedo; draft manuscript preparation: P. Fernandes, R. Tomás, E. Macedo, M.C. Coelho; project administration and supervision: M.C. Coelho. All authors reviewed the results and approved the final version of the manuscript.

6. REFERENCES

[1] Peter Mock, and Sonsoles Díaz. *Pathways to decarbonization: The European passenger car market, 2021–2035*. https://theicct.org/sites/default/files/publications/decarbonize-EU-PVs-may2021.pdf, Accessed June 9, 2021.

[2] Wang, Y., C. Hao, Y. Ge, L. Hao, J. Tan, X. Wang, P. Zhang, Y. Wang, W. Tian, Z. Lin, and J. Li. Fuel consumption and emission performance from light-duty

conventional/hybrid-electric vehicles over different cycles and real driving tests. *Fuel,* Vol. 278, 2020, p. 118340.
[3] ACAE. *New passenger cars by fuel type in the EU - First Quarter of 2021*. European Automobile Manufacturers' Association, Brussels, Belgium, https://www.acea.auto/fuel-pc/fuel-types-of-new-cars-battery-electric-5-7-hybrid-18-4-petrol-42-2-market-share-in-q1-2021/, Accessed June 9, 2021.
[4] Mi, C., and M. A. Masrur. *Hybrid Electric Vehicles: Principles and Applications with Practical Perspectives*. Wiley, 2017.
[5] Zhuang, W., S. Li, X. Zhang, D. Kum, Z. Song, G. Yin, and F. Ju. A survey of powertrain configuration studies on hybrid electric vehicles. *Applied Energy,* Vol. 262, 2020, p. 114553.
[6] Zhang, P., F. Yan, and C. Du. A comprehensive analysis of energy management strategies for hybrid electric vehicles based on bibliometrics. *Renewable and Sustainable Energy Reviews,* Vol. 48, 2015, pp. 88-104.
[7] Borthakur, S., and S. C. Subramanian. Design and optimization of a modified series hybrid electric vehicle powertrain. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering,* Vol. 233, No. 6, 2019, pp. 1419-1435.
[8] Wall, M. *Automotive Industry Outlook: Managing Volatility and Leveraging Opportunities in a Dynamic Market Environment - Global Propulsion Design Islands in 2031*. https://www.cargroup.org/wp-content/uploads/2019/02/Wall.pdf, Accessed June 9, 2021.
[9] Holmén, B. A., and K. M. Sentoff. Hybrid-Electric Passenger Car Carbon Dioxide and Fuel Consumption Benefits Based on Real-World Driving. *Environmental Science & Technology,* Vol. 49, No. 16, 2015, pp. 10199-10208.
[10] Huang, Y., N. C. Surawski, B. Organ, J. L. Zhou, O. H. H. Tang, and E. F. C. Chan. Fuel consumption and emissions performance under real driving: Comparison between hybrid and conventional vehicles. *Science of The Total Environment,* Vol. 659, 2019, pp. 275-282.
[11] Duarte, G. O., R. A. Varella, G. A. Gonçalves, and T. L. Farias. Effect of battery state of charge on fuel use and pollutant emissions of a full hybrid electric light duty vehicle. *Journal of Power Sources,* Vol. 246, 2014, pp. 377-386.
[12] Conger, M., and B. A. Holmén. Characterization of Real-World Particle Number Emissions during Reignition Events from a 2010 Light-Duty Hybrid Electric Vehicle. *Transportation Research Record,* Vol. 2503, No. 1, 2015, pp. 137-146.
[13] Robinson, M. K., and B. A. Holmén. Onboard, Real-World Second-by-Second Particle Number Emissions from 2010 Hybrid and Comparable Conventional Vehicles. *Transportation Research Record,* Vol. 2233, No. 1, 2011, pp. 63-71.
[14] Fernandes, P., R. Tomás, E. Ferreira, B. Bahmankhah, and M. C. Coelho. Driving aggressiveness in hybrid electric vehicles: Assessing the impact of driving volatility on emission rates. *Applied Energy,* Vol. 284, 2021, p. 116250.
[15] Wang, Y., J. Wang, C. Hao, X. Wang, Q. Li, J. Zhai, Y. Ge, L. Hao, and J. Tan. Characteristics of instantaneous particle number (PN) emissions from hybrid electric vehicles under the real-world driving conditions. *Fuel,* Vol. 286, 2021, p. 119466.
[16] Zhai, H., H. Christopher Frey, and N. M. Rouphail. Development of a modal emissions model for a hybrid electric vehicle. *Transportation Research Part D: Transport and Environment,* Vol. 16, No. 6, 2011, pp. 444-450.
[17] Alvarez, R., and M. Weilenmann. Effect of low ambient temperature on fuel consumption and pollutant and CO2 emissions of hybrid electric vehicles in real-world conditions. *Fuel,* Vol. 97, 2012, pp. 119-124.

[18] Alvarez, R., P. Schlienger, and M. Weilenmann. Effect of hybrid system battery performance on determining CO2 emissions of hybrid electric vehicles in real-world conditions. *Energy Policy,* Vol. 38, No. 11, 2010, pp. 6919-6925.
[19] Robinson, M. K., and B. A. Holmén. Hybrid-electric passenger car energy utilization and emissions: Relationships for real-world driving conditions that account for road grade. *Science of The Total Environment,* Vol. 738, 2020, p. 139692.
[20] Sullivan, J. L., and K. Sentoff. Identifying Roadway Physical Characteristics That Contribute to Emissions Differences between Hybrid and Conventional Vehicles. *Transportation Research Record,* Vol. 2674, No. 10, 2020, pp. 599-613.
[21] Liu, J., A. Khattak, and X. Wang. The role of alternative fuel vehicles: Using behavioral and sensor data to model hierarchies in travel. *Transportation Research Part C: Emerging Technologies,* Vol. 55, 2015, pp. 379-392.
[22] Rios-Torres, J., J. Liu, and A. Khattak. 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,* Vol. 13, No. 2, 2019, pp. 123-137.
[23] Zhang, L., X. Zhao, and A. Khattak. A New Fuel Consumption Model Considering Vehicle's Speed, Acceleration and Jerk. Presented at Presented at the 99th Transportation Research Board Annual Meeting, 12-16 January, Washington D.C., U.S., 2020.
[24] Liu, J., X. Wang, and A. Khattak. Customizing driving cycles to support vehicle purchase and use decisions: Fuel economy estimation for alternative fuel vehicle users. *Transportation Research Part C: Emerging Technologies,* Vol. 67, 2016, pp. 280-298.
[25] Wang, X., A. J. Khattak, J. Liu, G. Masghati-Amoli, and S. Son. What is the level of volatility in instantaneous driving decisions? *Transportation Research Part C: Emerging Technologies,* Vol. 58, 2015, pp. 413-427.
[26] Ferreira, E., P. Fernandes, B. Bahmankhah, and M. C. Coelho. Micro-analysis of a single vehicle driving volatility and impacts on emissions for intercity corridors. *International Journal of Sustainable Transportation*, 2021, pp. 1-23.
[27] Eloisa, M. Two-Step Semidefinite Programming approach to clustering and dimensionality reduction. *Statistics, Optimization And Information Computing,* Vol. 3, No. 3, 2015.
[28] Vichi, M., and G. Saporta. Clustering and disjoint principal component analysis. *Comput. Stat. Data Anal.,* Vol. 53, No. 8, 2009, pp. 3194–3208.
[29] Macedo, E., and A. Freitas. The Alternating Least-Squares Algorithm for CDPCA.In, Springer International Publishing, Cham, 2015. pp. 173-191.
[30] Freitas, A., E. Macedo, and M. Vichi. An empirical comparison of two approaches for CDPCA in high-dimensional data. *Statistical Methods & Applications*, 2020.
[31] ACAE. *New passenger cars by fuel type in the EU - full-year 2020*. European Automobile Manufacturers' Association, Brussels, Belgium, https://www.acea.auto/fuel-pc/fuel-types-of-new-cars-electric-10-5-hybrid-11-9-petrol-47-5-market-share-full-year-2020/, Accessed June 16, 2021.
[32] 3DATX. *http://www.3datx.com/*, 3-dimensional data analysis (3DATX) Corporation.
[33] Sandhu, G., and H. Frey. Effects of Errors on Vehicle Emission Rates from Portable Emissions Measurement Systems. *Transportation Research Record: Journal of the Transportation Research Board,* Vol. 2340, 2013, pp. 10-19.
[34] Pascale, A., P. Fernandes, C. Guarnaccia, and M. C. Coelho. A study on vehicle Noise Emission Modelling: Correlation with air pollutant emissions, impact of kinematic variables and critical hotspots. *Science of The Total Environment,* Vol. 787, 2021, p. 147647.

[35] Wei, T., and H. C. Frey. Sensitivity of light duty vehicle tailpipe emission rates from simplified portable emission measurement systems to variation in engine volumetric efficiency. *Journal of the Air & Waste Management Association*, 2021, pp. 1-21.
[36] Yazdani Boroujeni, B., and H. C. Frey. Road grade quantification based on global positioning system data obtained from real-world vehicle fuel use and emissions measurements. *Atmospheric Environment,* Vol. 85, 2014, pp. 179-186.
[37] USEPA. *Methodology for developing modal emission rates for EPA's multi-scale motor vehicle & equipment emission system*, Ann Arbor, MI: Prepared by North Carolina State University for US Environmental Protection Agency; 286 p. Report No.: EPA420-R-02-027, 2002.
[38] EPA. *Section 86.144-94 - Calculations; exhaust emissions*. https://www.govinfo.gov/content/pkg/CFR-2012-title40-vol19/xml/CFR-2012-title40-vol19-sec86-144-94.xml, Accessed July 1, 2021.