

Article

A Methodological Framework for Inferring Energy-Related Operating States from Limited OBD Data: A Single-Trip Case Study of a PHEV

Michal Loman ^{1,*}, Branislav Šarkan ¹, Arkadiusz Małek ^{2,*}, Jacek Caban ³, Beata Martyna-Syroka ² and Katarzyna Piotrowska ³

¹ Department of Road and Urban Transport, Faculty of Operation and Economics of Transport and Communications, University of Žilina, Univerzitná 8215/1, 01026 Žilina, Slovakia; branislav.sarkan@fpedas.uniza.sk

² Department of Transportation and Informatics, WSEI University, 20-209 Lublin, Poland; beatamartyna96@gmail.com

³ Faculty of Mechanical Engineering, Lublin University of Technology, Nadbystrzycka 36, 20-618 Lublin, Poland; j.caban@pollub.pl (J.C.); k.piotrowska@pollub.pl (K.P.)

* Correspondence: michal.loman@fpedas.uniza.sk (M.L.); arkadiusz.malek@wsei.lublin.pl (A.M.)

Abstract

This paper presents a methodological framework for inferring energy-related operating states of plug-in hybrid electric vehicles (PHEVs) under conditions of limited and incomplete on-board diagnostic (OBD) data. The proposed approach is illustrated using a single short real-world urban trip recorded for one PHEV operating in electric mode. Unsupervised clustering based on k-means is applied in progressively expanded state spaces (3D–5D) to decompose the driving process into physically interpretable operating states, despite the absence of direct measurements of key variables such as regenerative braking power. Cluster validity indices, per-cluster silhouette values, temporal segmentation, and robustness checks are employed to support the interpretability and internal consistency of the results. The study demonstrates that even a single, non-representative OBD time series contains sufficient internal structure to recover meaningful energy-related information when appropriate state-space decomposition is applied. While no statistical generalization is intended, the results highlight the potential of the proposed framework for analyzing real-world vehicle operation under constrained data availability.

Keywords: plug-in hybrid electric vehicle (PHEV); unsupervised clustering; driving signatures; electric mode; renewable charging; photovoltaic carport; intelligent mobility; energy efficiency; emission reduction; sustainable transport



Academic Editors: Weiwei Yang and Yang Tian

Received: 12 November 2025

Revised: 12 December 2025

Accepted: 15 December 2025

Published: 17 December 2025

Citation: Loman, M.; Šarkan, B.; Małek, A.; Caban, J.; Martyna-Syroka, B.; Piotrowska, K. A Methodological Framework for Inferring Energy-Related Operating States from Limited OBD Data: A Single-Trip Case Study of a PHEV. *Vehicles* **2025**, *7*, 165. <https://doi.org/10.3390/vehicles7040165>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Plug-in hybrid electric vehicles (PHEVs) are increasingly operated in pure electric mode during short urban trips, which makes them a relevant subject for studies based on real-world driving data. However, the analysis of PHEV behavior in electric mode remains challenging due to limited access to high-resolution powertrain data and the constraints imposed by standard on-board diagnostics (OBD) interfaces [1]. As a result, many studies rely on laboratory measurements or aggregated indicators, which do not fully capture the temporal dynamics of real-world vehicle operation [2].

OBD-based datasets typically provide only a restricted set of variables, such as vehicle speed, estimated electric motor power, and battery state-of-charge, while direct measure-

ments of regenerative braking power or traction currents are often unavailable [3]. This limitation necessitates the use of data-driven and exploratory methods capable of extracting meaningful information from incomplete and indirect measurements. In this context, unsupervised learning techniques offer a pragmatic approach for identifying recurring operating patterns without requiring explicit physical models or labeled data [4].

Among unsupervised methods, k-means clustering is frequently used for partitioning multivariate time series into groups of similar operating states [5]. Its computational simplicity and transparency make it suitable for exploratory analyses, particularly when the objective is not prediction or classification but rather the identification of internally consistent data groupings. Nevertheless, the application of k-means to real-world vehicle OBD data requires careful methodological framing, especially when the available dataset is limited in size and scope [6].

The present study is therefore positioned as a methodological case study that demonstrates how unsupervised k-means clustering can be applied to a single real-world OBD time series recorded during a short urban trip of a PHEV operating exclusively in electric mode [7,8]. The aim of the study is not to derive generalizable conclusions about PHEV driving behavior but to illustrate a transparent and reproducible analytical workflow for extracting interpretable operating states from limited OBD data.

Specifically, the study investigates whether clustering performed in a low-dimensional feature space—comprising electric motor power at the wheels, vehicle speed, and longitudinal acceleration—can produce internally consistent and physically interpretable groupings of driving states [9]. By deliberately limiting the scope to a single illustrative dataset, the analysis focuses on methodological clarity, separation between data processing and interpretation, and explicit reporting of clustering outcomes [10].

This narrowly defined scope allows the proposed approach to be evaluated as a proof-of-feasibility for exploratory analysis of vehicle operating states under real-world constraints [11]. The methodology presented in this work may serve as a reference framework for future studies involving larger datasets, multiple vehicles, or extended experimental campaigns, where more comprehensive validation and generalization can be pursued.

Plug-in hybrid vehicles, which operate in purely electric mode, are an attractive alternative for commercial fleets and public services. In applications such as the police, city guards, and municipal services, PHEVs allow for most daily tasks to be performed emission-free, which is particularly important in densely populated urban areas. Their quiet operation in electric mode increases the effectiveness of patrol and intervention operations, especially at night. For transport services such as taxis and courier companies, the ability to charge the vehicle from the grid or a photovoltaic carport translates into real operating savings. PHEVs can operate in zero-emission mode during deliveries in restricted traffic areas while using the combustion engine on out-of-town routes, maintaining full operational flexibility.

Although numerous studies investigate the operation of hybrid and plug-in hybrid vehicles, they primarily focus on fuel consumption, combined hybrid modes, or laboratory driving cycles [12]. There is a lack of research providing a quantitative and multidimensional characterization of PHEV behavior exclusively in electric mode under real-world urban conditions [13]. Existing approaches rarely integrate kinematic data with topographic and energy-related variables, and they typically do not infer regenerative braking states when negative power is not available from the OBD system [14,15].

Therefore, the novelty of this study is expressed through the following contributions:

1. Development of operational signatures of a PHEV driving in pure electric mode using unsupervised clustering applied to real-world OBD data [16].

2. Use of a multidimensional feature space (power, speed, acceleration, altitude, SOC), allowing for a physically interpretable decomposition of operating states.
3. Introduction of an indirect method for identifying regenerative braking phases using kinematic–topographic relations, overcoming the limitations of OBD systems.
4. Demonstration of the influence of urban topography on the electric drivetrain’s operating states, an aspect rarely addressed in PHEV research [17].
5. Construction of a clustering-based methodology that can be directly used for eco-driving assessment, route optimization, and fleet-level energy efficiency analysis [18].

These contributions clearly distinguish the present study from previous work and fill an identified research gap concerning real-world electric-mode behavior of plug-in hybrids.

The objective of this study is to develop and demonstrate a methodological framework for inferring physically interpretable energy-related operating states of a plug-in hybrid electric vehicle from limited and incomplete real-world OBD data. The framework is illustrated using a single short driving realization in electric mode and is based on unsupervised clustering applied in progressively expanded state spaces. The proposed approach aims to show how meaningful information about vehicle operating states can be recovered despite restricted data availability, rather than to provide statistically generalizable performance indicators or population-level driving signatures.

2. Materials and Methods

2.1. Test Vehicle and Data Acquisition

The experimental investigation was conducted using a Kia Ceed SW Plug-in Hybrid Electric Vehicle (PHEV) equipped with a 1.6 L GDI gasoline engine and an 8.9 kWh lithium-ion polymer traction battery. The nominal system voltage of the electric subsystem is 360 V, and the combined power output of the hybrid drivetrain is 104 kW. All measurements were performed under urban driving conditions, with the vehicle operating exclusively in electric mode (EV). During the experiment, the internal combustion engine remained inactive, ensuring zero local emissions and full electric propulsion. The Kia Ceed SW Plug-in Hybrid test vehicle was manufactured at the Kia Motors Slovakia plant in Žilina, which provides an additional context for research cooperation with the local university.

Vehicle operating data were acquired via the On-Board Diagnostics (OBD-II) interface using an OBDLink LX Bluetooth adapter connected to the Torque Pro mobile application (version 1.12.98). The sampling frequency was set to 1 Hz, resulting in approximately 3600 data records per hour of driving.

The analyzed dataset corresponds to one continuous real-world driving sequence with a duration of approximately 14 min and a total distance of approximately 8 km. The vehicle was driven by a single driver under typical urban traffic conditions. No attempt was made to ensure statistical representativeness of the trip; instead, the dataset was intentionally treated as an illustrative case study.

Vehicle operating data were acquired using the standard on-board diagnostics (OBD) interface. The OBD system provides indirect access to selected vehicle parameters at a limited sampling rate, without direct measurements of high-voltage currents or regenerative braking power. All data were recorded continuously for the entire duration of the trip.

2.2. Available Variables and Feature Selection

The raw OBD dataset included vehicle speed, estimated electric motor power at the wheels, battery state-of-charge (SOC), and auxiliary signals related to vehicle operation. Due to the known limitations of OBD systems, variables such as direct traction current or regenerative braking power were not available.

For the purpose of clustering, a deliberately low-dimensional feature space was selected, consisting of the following three variables:

- electric motor power at the wheels (kW),
- vehicle speed (km/h),
- longitudinal acceleration (m/s^2).

Longitudinal acceleration was computed numerically as the first-order time derivative of vehicle speed [19]. The selected feature set was chosen to reflect the dynamic operating state of the vehicle while maintaining transparency and minimizing redundancy. Variables related to topography or battery state were not included in the primary clustering analysis to preserve methodological clarity and limit the dimensionality of the feature space.

2.3. Data Preprocessing and Scaling

Prior to clustering, all selected variables were inspected for missing values and obvious recording artifacts. No data imputation was applied. The time series was synchronized using the original OBD timestamps.

To ensure comparable scaling of variables with different physical units, standardization was applied using the StandardScaler method, which transforms each variable to zero mean and unit variance. This choice avoids dominance of variables with larger numerical ranges and is commonly used in distance-based clustering methods.

Data processing and clustering workflow applied in the study is presented in Figure 1. Raw time-series data acquired via the OBD interface are transformed through feature selection and preprocessing steps before unsupervised k-means clustering. The resulting cluster labels and centroids are subsequently used for result representation.

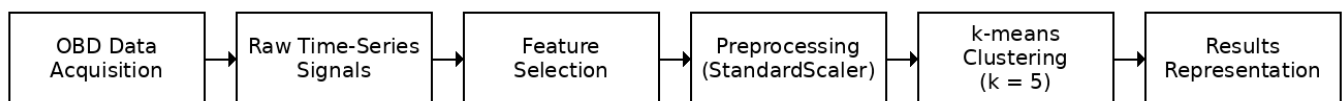


Figure 1. Workflow of data acquisition and analysis.

2.4. Unsupervised Clustering Procedure

Unsupervised clustering was performed using the k-means algorithm. The method partitions the dataset into k clusters by minimizing the within-cluster sum of squared Euclidean distances [20]. The algorithm was implemented with a fixed random initialization seed to ensure reproducibility [21].

Based on exploratory evaluation and consistency considerations, the number of clusters was set to $k = 5$. This value was selected to provide a balance between cluster separation and interpretability of the resulting operating states [22]. No attempt was made to optimize k as a global model parameter, as the objective of the study is methodological demonstration rather than model selection [23].

2.5. Output Representation and Consistency Checks

The clustering output consists of cluster labels assigned to each time step and the corresponding cluster centroids in the standardized feature space. Results are presented using:

- temporal plots showing vehicle speed over time with cluster assignments,
- relative frequencies of occurrence of individual clusters,
- centroid tables summarizing average feature values for each cluster.

These representations are intended to support internal consistency assessment and facilitate transparent interpretation of clustering outcomes.

2.6. Experimental Protocol and Repeatability

The repeatability of the experimental procedure was ensured primarily through the fixed sampling frequency of 1 Hz, which provided a uniform temporal resolution of all measured variables throughout the entire driving cycle. The OBDLink LX interface and Torque Pro application were initialized in an identical manner, guaranteeing consistent data retrieval and eliminating variability resulting from communication latency or sensor activation order. Before applying k-means clustering, all features were standardized using the StandardScaler method, which ensured reproducible scaling and prevented differences in variable magnitude from affecting cluster boundaries. To ensure full repeatability of the unsupervised clustering process, the random centroid initialization function was fixed, eliminating randomness in the starting conditions of the algorithm. As a result, repeated executions of the k-means model produced identical cluster assignments, which was confirmed by achieving an Adjusted Rand Index (ARI) of 1.00 between independent runs. This combination of stable data acquisition and deterministic clustering ensured that the obtained operating-state decomposition is fully repeatable and not affected by stochastic variation in the computational workflow.

3. Results

3.1. Preparing the Vehicle for Testing

Before commencing road testing, it was necessary to properly prepare the vehicle to ensure the reliability and repeatability of measurements. A technical inspection was performed, including inspection of basic systems: braking, steering, tires, lighting, and the condition of electrical connections. The vehicle under test was completely new, with only approximately 500 km on the odometer, ensuring no wear of mechanical components and proper operation of the control systems. Before testing, the traction battery was charged to approximately 50% of its SoC to simulate typical urban driving conditions. This vehicle preparation ensured stable operating parameters and eliminated the influence of random factors on the measurement results. A view of the research vehicle while charging the batteries from the carport is shown in Figure 2.



Figure 2. Charging the research vehicle from the photovoltaic carport.

3.2. Road Driving Course in Urban Conditions

The road test in urban conditions is shown in Figure 3. The road test lasted 14 min. During this time, a distance of 8.39 km was covered at a maximum speed of 81.36 km/h. The PHEV was driven in electric mode only.

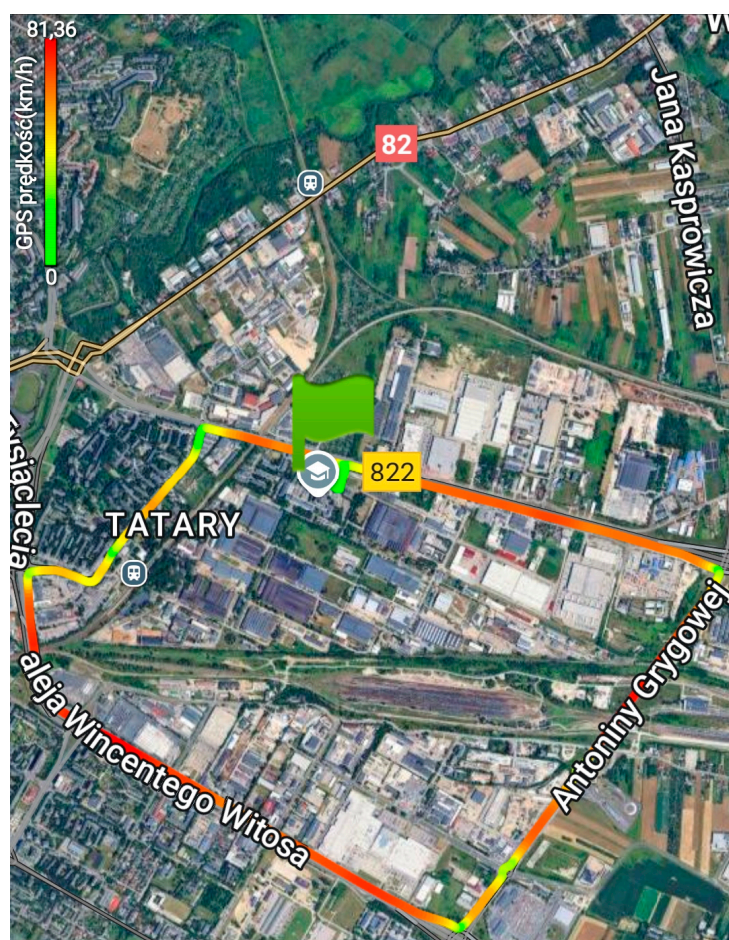


Figure 3. Route in urban conditions presented in the Torque Pro mobile application.

The recorded time series of driving speed and altitude above sea level (Figure 4) reflect typical traffic conditions in a city with diverse terrain, such as Lublin. Analysis of the speed curve indicates frequent changes in driving speed, characteristic of urban traffic with numerous intersections, traffic lights, and pedestrian traffic. The altitude values show clear terrain fluctuations, confirming the city's location in the Lublin Upland, where the terrain is undulating. During the drive, cyclical alternating sections of ascent and descent were observed, which translates into changes in the load on the vehicle's drive system. The altitude characteristics resemble the profile of a city situated on hills, similar to Rome, with numerous valleys and elevations. Such conditions favor the frequent use of energy recuperation in hybrid vehicles. Analysis of both time series allows us to conclude that Lublin's topography significantly affects the energy profile of driving and the efficiency of the electric drive.

Time series of electric motor power and hybrid battery charge levels (Figure 5) provide a detailed picture of the vehicle's drive system operation in electric mode. The electric motor's power varies dynamically in response to driver maneuvers, road conditions, and the terrain characteristic of Lublin. Power values remain only in the positive range due to limitations of the OBD system, which does not directly record regenerative braking power. Nevertheless, the data indirectly reveal moments of energy recovery through small increases in battery charge after power drop phases. Power fluctuations ranging from a few to several dozen kilowatts indicate varying drive load during city driving, including frequent acceleration and deceleration. The state-of-charge (SOC) value remains relatively stable, demonstrating the effective operation of the energy management system and protection of the battery against excessive discharge. The relationship between instantaneous

power and SOC indicates that the control system is maintaining a balance between energy demand and available battery capacity. Analysis of this data confirms that the vehicle exhibits typical characteristics of a plug-in hybrid in urban conditions—short acceleration cycles, smooth acceleration, and limited top speeds. The series characteristics reflect the harmonious interaction between the electric drive and the energy management system. The obtained results provide a basis for further analysis of driving efficiency and an assessment of the vehicle's energy potential in real-world driving conditions.

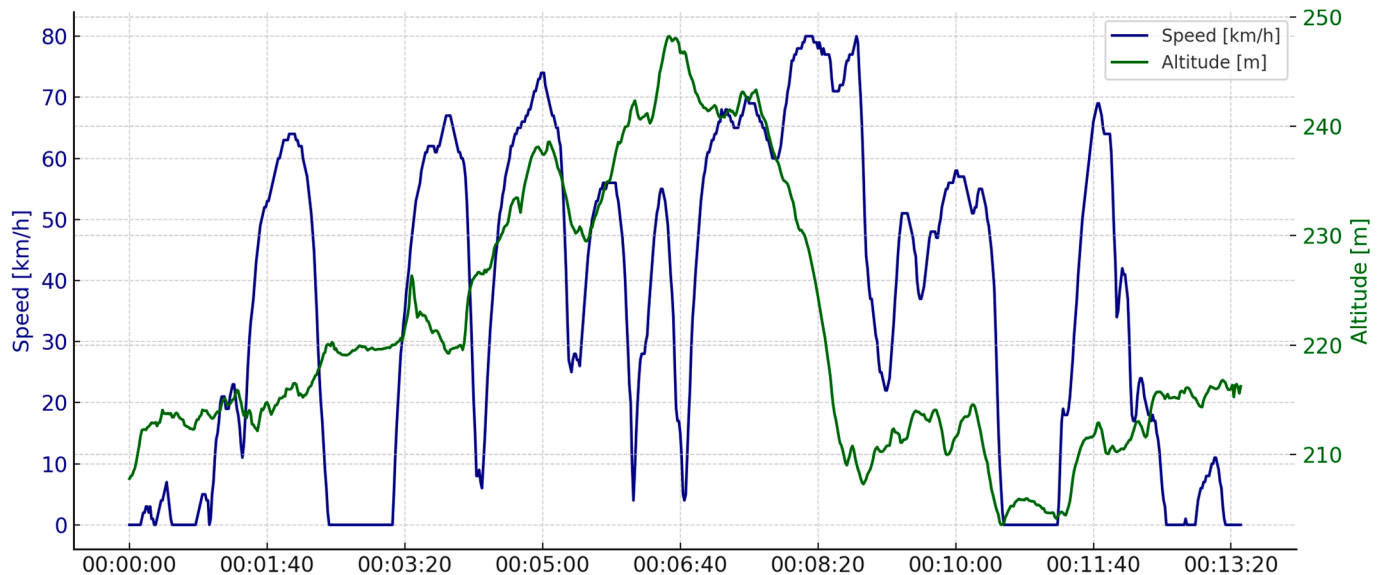


Figure 4. Time series of driving speed and altitude above sea level during a road test in urban conditions.

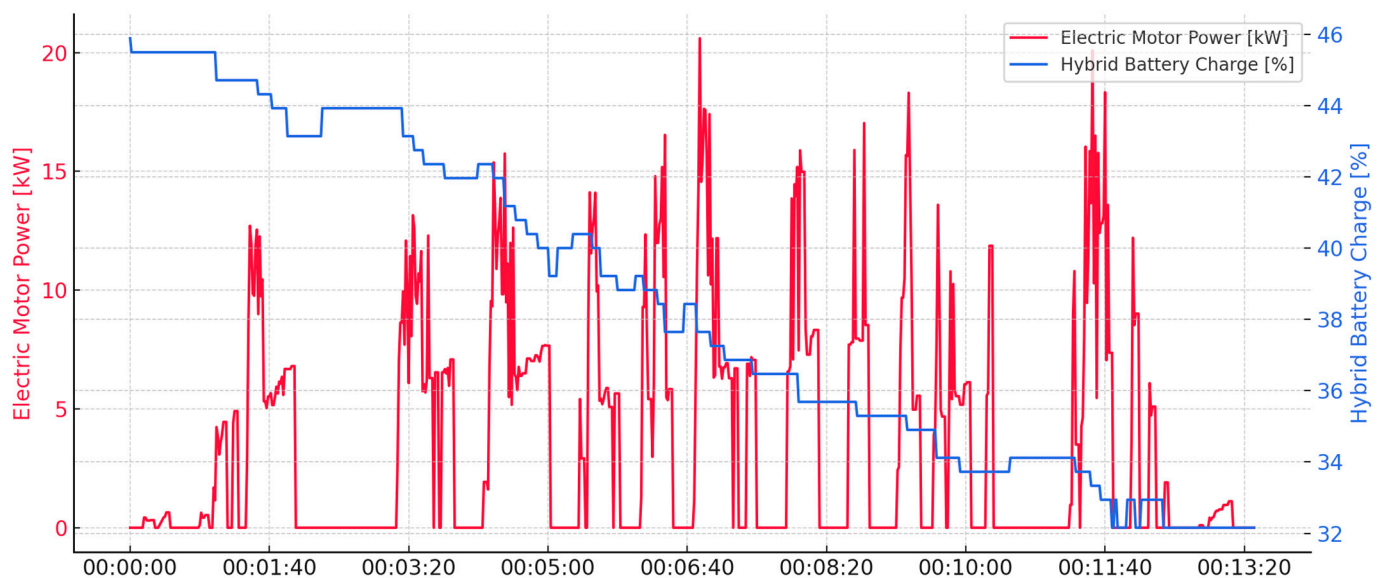


Figure 5. Time series of the power consumed by the electric motor and the state of charge of the traction battery.

3.3. Additional Calculations

Based on previous experience, the authors decided to expand the list of recorded parameters to include vehicle acceleration. The calculation was based on the driving speed

recorded during the tests. The time series of vehicle speed and acceleration is presented in the graph in Figure 6.

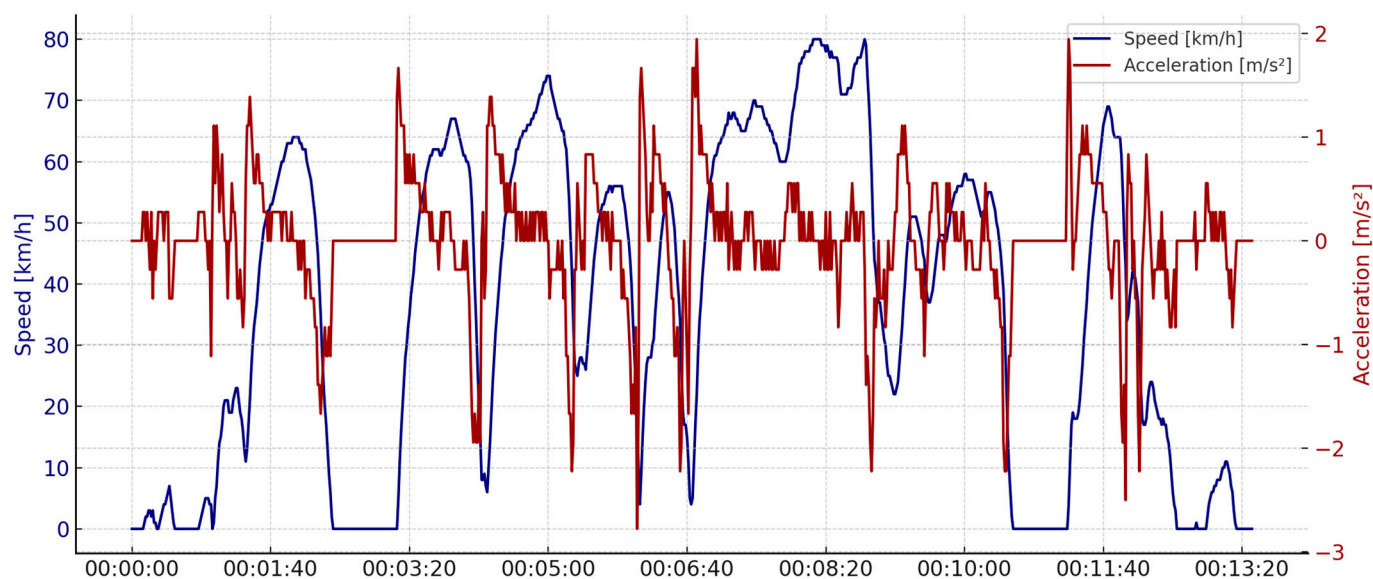


Figure 6. Time series of vehicle speed and acceleration.

The speed and acceleration time series reflect typical urban driving dynamics, characterized by frequent changes in pace and short acceleration and braking cycles. Speed values exhibit numerous fluctuations, indicating a diverse traffic pattern—from smooth driving to stop-and-go traffic. Acceleration varies widely, with values close to zero dominating, indicating frequent constant speed or slow changes. High, short-lived positive peaks correspond to intense acceleration phases, while negative acceleration values indicate moments of deceleration or braking. This characteristic confirms that the vehicle was operating in typical urban traffic conditions with high speed variability.

3.4. Measurement Data Processing—Unsupervised Clustering

The authors used unsupervised clustering in a five-dimensional space covering the following driving parameters:

- Hybrid Battery Charge (%)
- Electric motor power at the wheels (kW)
- OBD Speed (km/h)
- Acceleration (m/s^2)
- Height above sea level (m)

However, this was not achieved immediately. The research began with a preliminary determination of the significance of individual measured state parameters. In further analyses, the authors will employ a dual approach involving unsupervised clustering in a variable-dimensional space, along with a simultaneous assessment of clustering quality in the form of expert validation of the obtained results.

3.4.1. Unsupervised Clustering in 3-Dimensional Space

In the context of identifying the operating states of a hybrid vehicle in electric mode, three parameters are most important:

- Electric motor power at the wheels (kW)—directly describes the load level and energy flow,

- Acceleration (m/s^2)—reflects the dynamics of movement and allows for the distinction between acceleration and deceleration (an indirect indicator of recuperation),
- OBD Speed (km/h)—determines the driving phase (maneuvering, city traffic, smooth driving).

The Hybrid Battery Charge (SOC) and height above sea level parameters serve as auxiliary parameters:

- SOC changes slowly and to a small extent, therefore not contributing much variability to the feature space, but allows for the interpretation of the energy balance (whether the battery is discharging or charging).
- Height above sea level is a contextual parameter, useful for analyzing the impact of topography (e.g., Lublin—a city on hills) but not necessary for the division of clusters describing driving style and drive dynamics.

For operational analysis of electric drive operation, power, speed, and acceleration are the most important parameters. It is only worth retaining the SOC and height parameters if the purpose of the analysis is to assess energy efficiency or terrain impact; otherwise, they may be treated as secondary or redundant in clustering.

For a 3-dimensional state space, it is logical to begin clustering by dividing it into three clusters. However, considering the application of this technique to the analysis of vehicle road tests, the immediate conclusion is that three clusters are insufficient. The authors decided to increase the number of clusters until a cluster responsible for regenerative braking was identified. This cluster was identified only after decomposing the analyzed 3-dimensional state space into five clusters. Simultaneously, the authors calculated the silhouette index, which is used to assess the clustering quality. The results of these calculations are presented in Figure 7. The presented data indicate that increasing the number of clusters from $k = 3$ to $k = 6$ translates almost linearly into an increase in the silhouette index. However, increasing the number of clusters to $k = 7$ results in a small increase in the silhouette index.

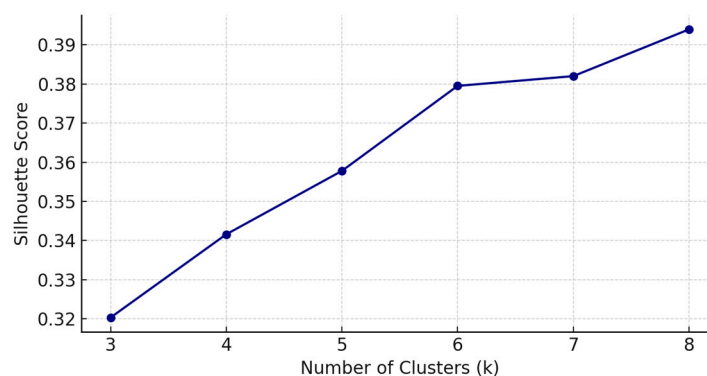


Figure 7. Silhouette score calculation results.

To ensure the robustness of the selected number of clusters, three commonly used internal cluster validity indices were evaluated for k ranging from 3 to 8:

1. Silhouette Coefficient (SC)—higher values indicate better cohesion and separation.
2. Davies–Bouldin Index (DBI)—lower values reflect better compactness and inter-cluster separation.
3. Calinski–Harabasz Index (CHI)—higher values indicate better variance ratio between clusters.

The comparative results showed that all three indices exhibit favorable values in the region of $k = 4$ – 6 . Specifically, the silhouette coefficient increases up to $k = 5$ and

stabilizes thereafter. The Davies–Bouldin Index reaches its lowest value at $k = 5$, indicating maximal cluster separation. The Calinski–Harabasz Index also shows a local maximum at $k = 5$, confirming that this configuration provides the best ratio of between-cluster to within-cluster dispersion.

Overall, the convergence of all three indices supports $k = 5$ as the optimal and most stable clustering configuration, offering a balance between mathematical validity and physical interpretability of driving states. The authors decided to begin their research by clustering the 3-dimensional state space into five clusters and then into six clusters. The choice of the number of clusters will be justified depending on the finding of a physical interpretation for the sixth cluster

The results of clustering in a 3-state space divided into 5 states (clusters) using the k-Means algorithm are presented in the graph in Figure 8.

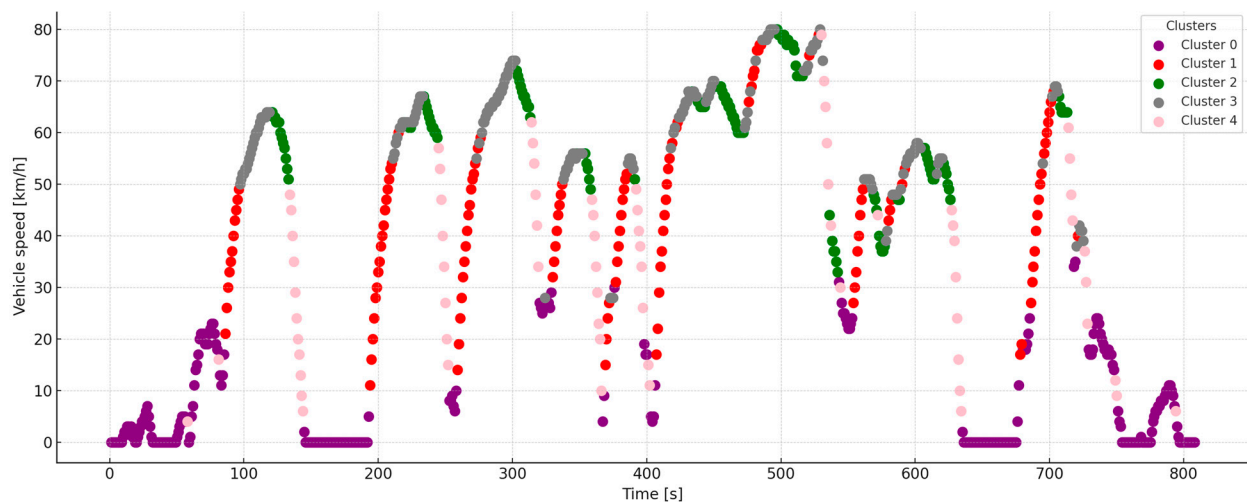


Figure 8. Results of clustering in a 3-state space divided into 5 states (clusters) using the K-Means algorithm.

Then, the physical interpretation of the clusters was performed based on expert validation.

Cluster 0 (purple) represents slow driving or maneuvering phases with very low electric motor power. The average speed in this cluster is around 6.5 km/h, corresponding to situations typical of driving in traffic jams, approaching intersections, parking, or maneuvering in tight spaces. The acceleration value is close to zero ($\approx 0.08 \text{ m/s}^2$), indicating small speed changes and no dynamic maneuvers. The electric motor power remains very low ($\approx 0.55 \text{ kW}$), indicating minimal stress on the drivetrain and operation in a high energy efficiency range. In these conditions, the vehicle uses electric propulsion in the most economical way, without engaging the combustion engine. This cluster can therefore be considered representative of quiet, low-energy urban driving, typical of PHEVs in congested traffic or during precise maneuvers.

Cluster 1 (red) corresponds to dynamic acceleration phases in urban conditions, where the electric system delivers high power to the wheels. The average engine power in this cluster is approximately 12.2 kW, indicating significant energy demand, typical of intense acceleration or hill climbing. This is accompanied by a positive acceleration of approximately 0.69 m/s^2 , indicating a significant increase in speed. The average vehicle speed in this cluster is approximately 43 km/h, suggesting the vehicle is accelerating from low to medium speeds, for example, when merging into traffic or exiting built-up areas. From an energy perspective, this cluster represents phases of high traction battery load, where electric energy is intensively used for propulsion. Such episodes are often preceded

by cluster 3 (braking), creating a natural acceleration–deceleration cycle typical of urban driving. Therefore, it can be concluded that cluster 0 reflects the active operation of the electric drive, which determines energy consumption and the vehicle’s dynamic profile in EV mode [24].

Cluster 2 (green) represents phases of steady driving or very gentle deceleration at moderate speeds and minimal power consumption. The average vehicle speed in this cluster is approximately 61 km/h, indicating smooth driving in rural traffic or on main arterial roads with few stops. The average electric motor power is close to zero (≈ 0 kW), and the acceleration is -0.19 m/s^2 , meaning the vehicle either drives with a slight decrease in speed or maintains it steadily, with periodic moments of releasing the accelerator pedal. In this situation, the electric drive switches to a low-energy phase, reducing power consumption and maintaining high energy efficiency. This cluster can also encompass short driving sections with minimal kinetic energy recuperation, particularly during gentle descents. From an operational perspective, it corresponds to the most economical electric driving style, in which the drive system maintains speed without sudden torque changes. The green cluster can therefore be considered characteristic of stable, energy-efficient driving in EV mode, typical of sections with constant speed and low traffic.

Cluster 3 (gray) corresponds to smooth driving at higher speeds with moderate electric drive load. Average values ($\sim 61 \text{ km/h}$, $\sim 6.6 \text{ kW}$, $+0.10 \text{ m/s}^2$) indicate stabilized traffic conditions—through roads, bypasses, or major urban arteries, where there is no intense acceleration or braking [25]. A slight, positive acceleration suggests maintaining speed rather than dynamically increasing it. From an energy perspective, this is a high-efficiency regime, as constant speed and moderate power limit losses are associated with frequent torque changes. In this cluster, the SOC typically changes slowly, and the energy balance is predictable, which favors range planning in EV mode. In Lublin’s topography, this condition occurs primarily on gentle, level sections, possibly on slight inclines that do not require significant power peaks. Operationally, cluster 4 can be treated as “economical cruising”—beneficial for comfort, noise, and energy consumption.

Cluster 4 (pink) very likely represents regenerative braking phases, albeit with a caveat due to OBD data limitations.

The reasoning is as follows:

- The average acceleration in this cluster is approximately -1.45 m/s^2 , which clearly indicates deceleration, i.e., the braking phase.
- At the same time, the electric motor’s power is very low (0.12 kW), meaning it does not provide traction power. In reality, it could operate in generator mode during this phase, but the OBD does not record negative power values.
- The average speed of 34 km/h confirms that the vehicle is not stationary but is decelerating smoothly—typical conditions in which the hybrid system switches to energy recovery mode [26].

The timeline shows that this cluster is short-lived and occurs after high-power episodes (cluster 0), which corresponds to the natural sequence: acceleration $\rightarrow$ deceleration with recuperation. Cluster 3 can be identified with high probability with regenerative braking phases, even though the OBD measurement does not record this process directly—its presence is indirectly inferred from the characteristic combination of low power, negative acceleration and speed drop.

The relative frequency of each cluster was then calculated. The results are presented in Figure 9.

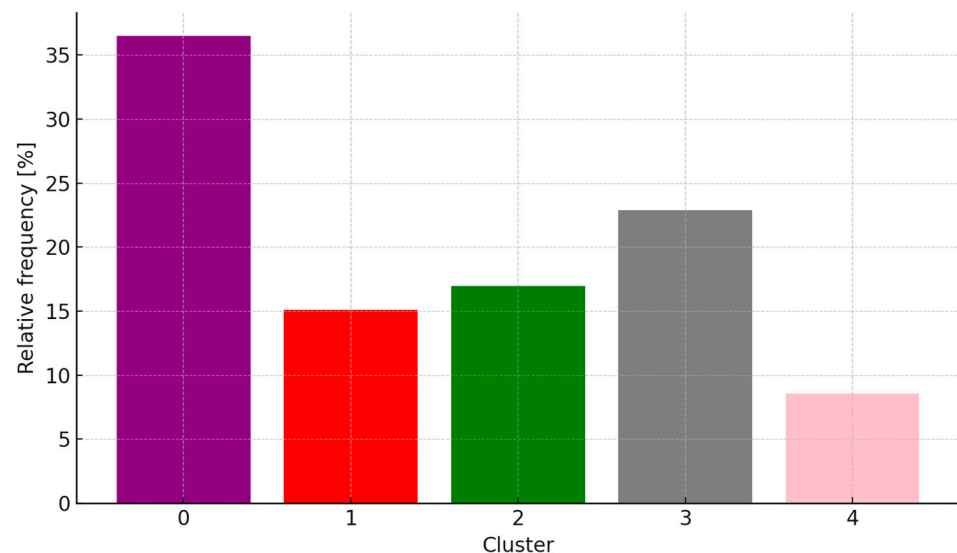


Figure 9. Relative frequencies of occurrence of individual 5 clusters.

In contrast to the technical interpretation above, this section discusses how frequently each cluster occurred during the trip and how these frequencies reflect the overall structure of urban driving in hilly terrain. The graph in Figure 9 shows the relative frequency of five clusters from the K-Means analysis ($k = 5$): Cluster 0—purple (36.51%) corresponds to slow driving/maneuvering with low power; Cluster 3—gray (22.90%) describes smooth driving with a constant, higher speed and moderate power; Cluster 2—green (16.96%) represents steady driving or very gentle deceleration with minimal energy consumption [27]; Cluster 1—red (15.10%) represents dynamic acceleration with a pronounced positive acceleration and increased power; Cluster 4—gray (8.54%) reflects episodes of more intense deceleration/braking, in which we infer recuperation (indirectly, based on negative acceleration and low power reported by the OBD). Interpretively, the dominance of Cluster 0—purple (36.5%) is typical of urban traffic: numerous stops, traffic light runs, traffic jams, and parking maneuvers favor long stretches of low speed and low power. The significant share of Cluster 3—gray (22.9%) indicates that, despite urban conditions, there were longer, smoother journeys (e.g., main arteries, ring roads), which were energy-efficient in EV mode. Cluster 2—green (17.0%) and Cluster 1—red (15.1%) clusters form a natural “cruise $\leftrightarrow$ acceleration” sequence: green corresponds to economical speed maintenance with minimal power, and purple to intense speed increases requiring higher battery consumption [28]. The rarest Cluster 4—gray (8.5%) appears in short episodes, which confirms the punctual nature of braking in city traffic; in Lublin’s hilly terrain, some of these episodes may be related to descents, but the OBD does not directly register negative recuperative power—we infer it indirectly from kinematics (negative acceleration). In the context of the topography of the Lublin Upland, such a cluster distribution is consistent: numerous hills force more frequent purple episodes (acceleration/uphills), interspersed with gray ones (slowing downhills), and smooth sections reward pink and green (smooth, efficient driving). Overall, the cluster picture confirms that city driving in Lublin combines high speed variability with a distinct topographic component, which favors “acceleration–deceleration” cycles and periodic recuperation, while maintaining longer, economical sections of smooth driving.

The values of the centroids are presented both in standardized space (Table 1) and in physical units (Table 2), allowing for an unambiguous interpretation of the characteristic operating states of the vehicle. Subsequently, a multidimensional validation of the clustering was performed, including silhouette analysis and temporal analysis for each cluster,

which enabled the assessment of cluster separability and the temporal stability of their occurrence. Finally, as part of the clustering quality assessment, a Robustness Check was conducted, confirming the repeatability of the cluster assignments through full consistency between independent runs of the algorithm.

Table 1. Centroids in standardized space.

Cluster	Speed OBD (z)	Power (z)	Acceleration (z)
0	−1.114	−0.638	0.081
1	0.273	1.831	0.685
2	0.932	−0.755	−0.193
3	0.929	0.642	0.101
4	−0.062	−0.729	−1.445

Table 2. Centroids in physical units.

Cluster	Speed OBD (z)	Power (z)	Acceleration (z)
0	6.45	0.55	0.081
1	43.43	12.17	0.685
2	60.99	≈0.00	−0.193
3	60.92	6.57	0.101
4	34.49	0.12	−1.445

The centroid tables showed five distinct driving modes: low-speed driving, dynamic acceleration, steady driving at higher speeds (two variants) and intensive braking phases with recuperation.

Validation was then performed based on silhouette analysis (Table 3). The global silhouette score was 0.512 (average), and the mean silhouette values for individual clusters are presented in Table 3.

Table 3. Average silhouette values for clusters.

Cluster	Mean Silhouette
0	0.603
1	0.306
2	0.561
3	0.519
4	0.368

Validation conclusions:

- Clusters 0, 2, and 3 are well separated (silhouette ~0.52–0.60).
- Clusters 1 and 4 are more diffuse, partially overlapping neighboring clusters (silhouette ~0.31–0.37).

A global $S \approx 0.51$ suggests reasonable, but not perfect, separation. This is typical for driving data with smooth transitions between modes. Clustering achieved an average silhouette coefficient of 0.51, indicating moderately good separation between clusters while maintaining the continuous nature of the driving data.

In the next step of the clustering evaluation, a temporal analysis was performed (Table 4). Continuous segments were identified from the temporal course of clusters, where the cluster number did not change. A summary for each cluster is presented in Table 4.

Table 4. Summary of temporal analysis for each cluster.

Cluster	Number of Segments	Average Duration [s]	Maximum Duration [s]
0	19	15.5	57
1	26	4.7	17
2	20	6.8	22
3	38	4.9	26
4	18	3.8	11

Interpretation resulting from temporal analysis:

- Cluster 0 (low speeds)—longer, less frequently interrupted periods (up to 57 s); this corresponds to driving at very low speeds/standing still.
- Cluster 1 (rapid acceleration)—short episodes (~5 s on average), typical of acceleration phases.
- Clusters 2 and 3 (steady driving at higher speeds)—average segments of 5–7 s, sometimes longer (up to 22–26 s).
- Cluster 4 (braking/recuperation)—the shortest episodes, ~4 s, consistent with short phases of intense deceleration.

This combination can be called temporal stability and driving pattern duration for the purposes of driving style analysis or signature determination.

Temporal analysis showed that acceleration and braking phases are short-lived (on average 4–5 s), while low-speed driving and steady driving at higher speeds occur in longer, stable sections (up to 57 s).

A Robustness Check was then performed for the following assumptions:

- Base model: k-Means ($k = 5$, $n_init = 10$, $random_state = 0$)
- Alternative model: k-Means ($k = 5$, $n_init = 5$, $random_state = 1$)
- Goodness of fit: Adjusted Rand Index (ARI)

The obtained result was $ARI = 1.00$. This means that changing the initialization (different random generator seed) resulted in identical cluster allocation ($ARI = 1.0$). This result indicates very high robustness of the clustering solution for the assumed data and K-means parameters.

Clustering of five driving modes, after standardizing three parameters (speed, power, and acceleration), allowed the identification of coherent and interpretable groups corresponding to the actual dynamic states of the vehicle. A global silhouette score of 0.51 indicates moderately good cluster separation, which is typical for continuous data with fluid boundaries between driving states. Centroid analysis reveals clear differences between clusters, including dynamic acceleration, steady driving with varying power, and regenerative braking phases. The temporal stability of segments and the repeatability of assignments in the robustness check ($ARI = 1.0$) confirm the high reliability of the K-means algorithm. Overall, the clustering is effective, well-adapted to the data characteristics, and allows for a precise characterization of the hybrid vehicle's driving style.

In the next research step, the number of clusters was increased to $k = 6$. The results of this clustering are presented in the graph in Figure 10.

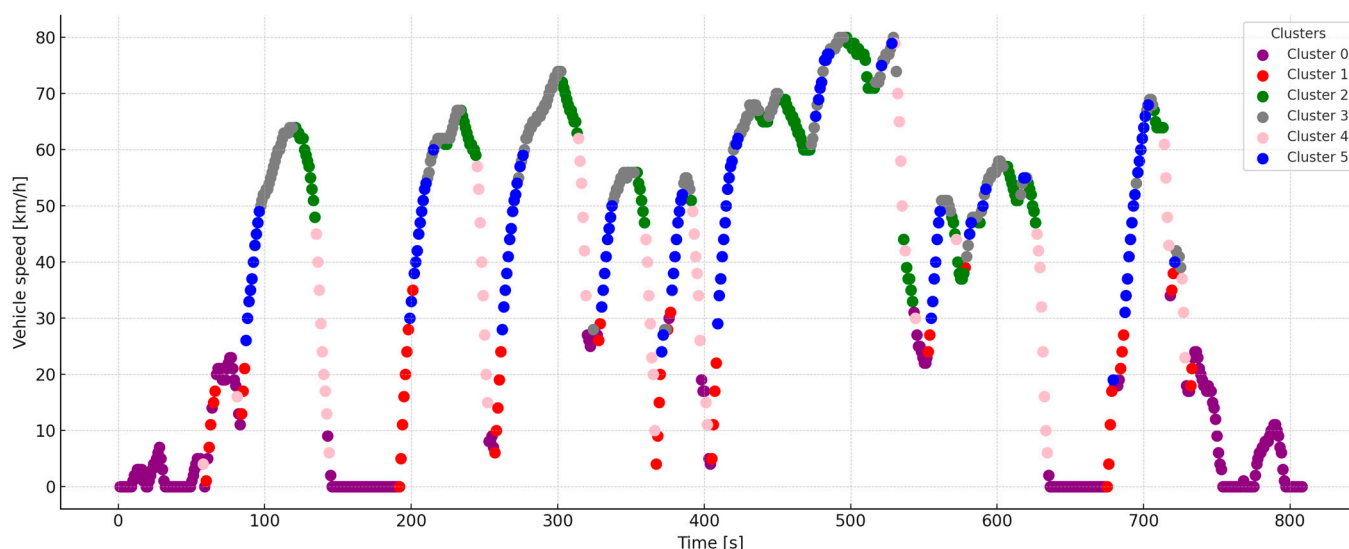


Figure 10. Results of clustering in a 3-state space divided into 6 states (clusters) using the K-Means algorithm.

Increasing the number of clusters to $k = 6$ provides some additional interpretive possibilities but also requires caution when analyzing the results. Compared to the previous division into five clusters, the new model allows for better differentiation of driving dynamics—particularly between gentle and strong acceleration and between light deceleration and full braking. In practice, this may mean that the clusters previously generally corresponding to “acceleration” and “braking” have been separated into two more precise operating states: gentle starting in city traffic and intense acceleration while climbing or overtaking.

From the perspective of PHEV energy analysis, this refinement can be valuable—it allows for better identification of when the electric drive is operating at optimal efficiency and when it is approaching its maximum load limits. On the other hand, the increase in classification quality (silhouette $\sim 0.38 \rightarrow 0.39$) is small, meaning that some of the new clusters may have limited interpretability or represent transient states rather than distinct operating modes. Increasing the number of clusters to six allows for a more detailed segmentation of driving style and driving behavior, which is useful for detailed studies (e.g., optimizing control strategies, analyzing recuperation efficiency). However, for synthetic or comparative presentations (e.g., as in this research paper), a five-cluster model remains more readable and sufficient, maintaining a balance between detail and interpretability.

3.4.2. Unsupervised Clustering in 4-Dimensional Space

Next, the authors used unsupervised clustering in a four-dimensional space covering the following driving parameters:

- Electric motor power at the wheels (kW)
- OBD Speed (km/h)
- Acceleration (m/s^2)
- Height above sea level (m)

This analysis takes into account an additional (fourth) state space parameter in the form of Height above sea level (m)→altitude (Figure 11).

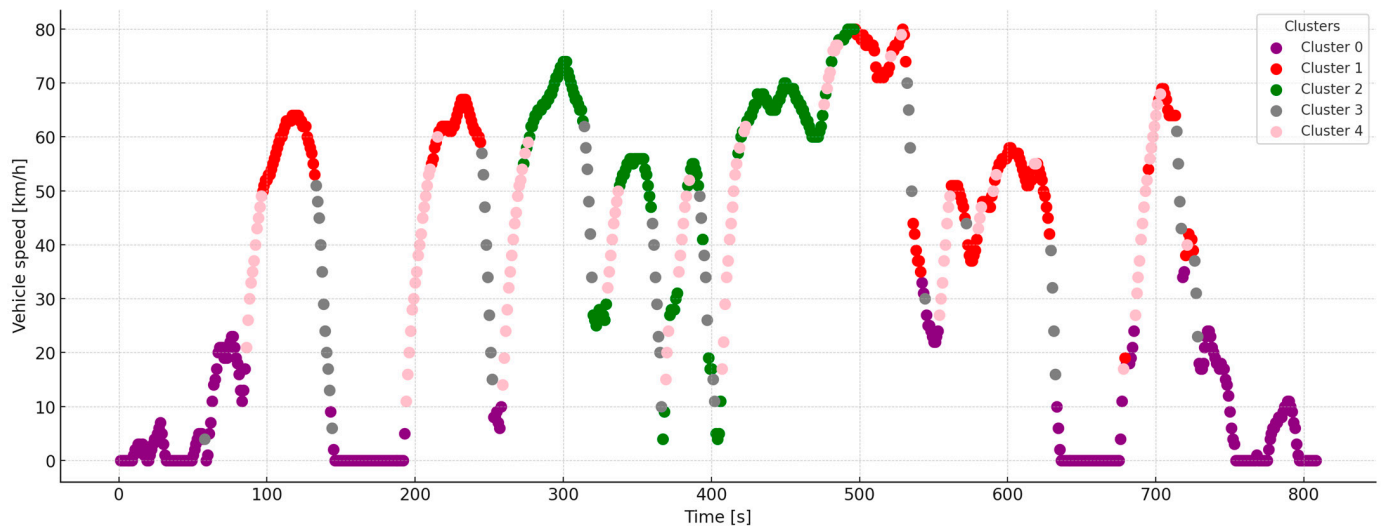


Figure 11. Results of clustering in a 4-state space divided into 5 states (clusters) using the K-Means algorithm.

The addition of a fourth parameter—height above sea level—to cluster analysis significantly enhanced the interpretation of vehicle behavior by introducing topographic context into the classification of driving states. Previous analyses based solely on power, speed, and acceleration described a dynamic state (e.g., acceleration, braking, steady driving) but did not distinguish between whether a given phenomenon occurred on an incline, descent, or flat section.

By adding height, it became possible to:

- link changes in speed and acceleration to terrain, which is particularly important in hilly Lublin, where topography significantly influences driving energy efficiency;
- identify episodes in which braking results from gravity descent rather than active driver input—such cases indicate potential points of intense recuperation;
- separate acceleration on uphill slopes (higher power consumption at lower speeds) from acceleration on flat terrain, allowing for more precise analysis of electric drive load;

This allows for the distinction of clusters with similar dynamic parameters but different terrain conditions, which increases classification precision without the need to increase the number of clusters.

In practice, this means that each driving state (e.g., acceleration, steady driving, braking) can now be described not only in kinematic terms but also in energy-topographic terms. Adding elevation reveals the relationships between terrain, energy consumption, and the frequency of recuperation, making the analysis more realistic and physically interpretable. In the case of a city with diverse terrain, such as Lublin, this fourth parameter is therefore crucial for fully understanding the driving characteristics and operating efficiency of PHEVs.

3.4.3. Unsupervised Clustering in 5-Dimensional Space

Finally, the authors used unsupervised clustering in a five-dimensional space covering the following driving parameters:

- Electric motor power at the wheels (kW)
- OBD Speed (km/h)
- Acceleration (m/s^2)
- Height above sea level (m)

- Hybrid Battery Charge (%)

In unsupervised clustering, the last (5th) state parameter, Hybrid Battery Charge (%), was taken into account. The clustering results are shown in Figure 12.

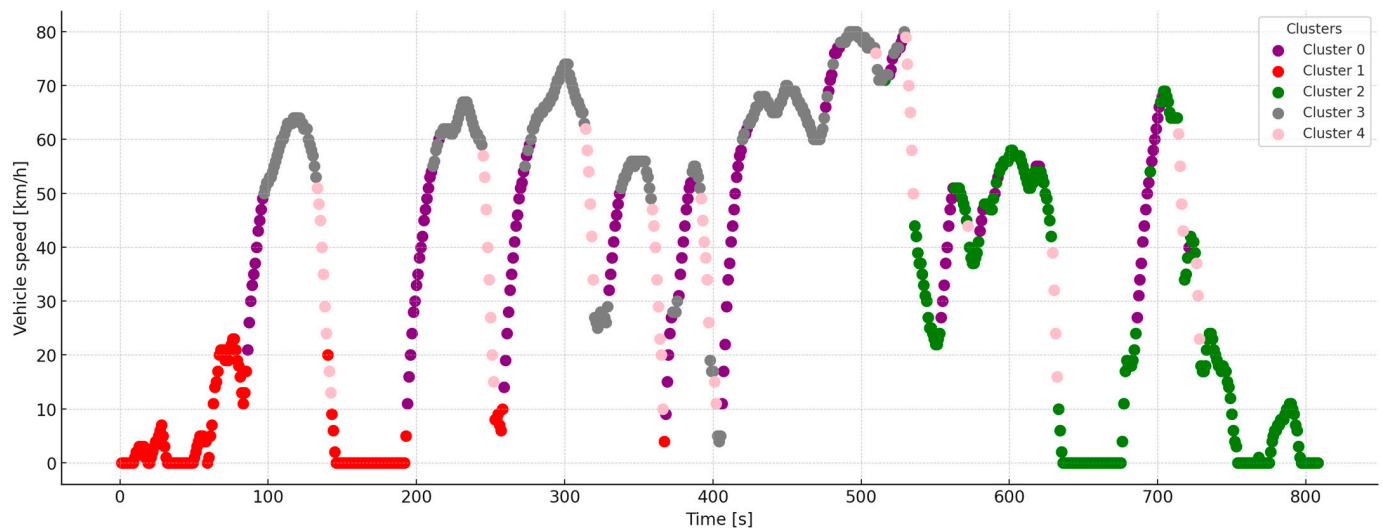


Figure 12. Results of clustering in a 5-state space divided into 5 states (clusters) using the K-Means algorithm.

The addition of a fifth parameter—the Hybrid Battery Charge (SOC)—to the cluster analysis allowed for the interpretation of results from the pure driving dynamics (speed, acceleration, power, topography) in a more comprehensive energy context of the hybrid system. While the previous four-parameter model effectively distinguished driving states dependent on topography (e.g., uphill, downhill, flat), it did not incorporate information about how the battery’s electric energy capacity changes during individual driving phases.

By incorporating SOC, it became possible to:

1. Distinguish between states that appear dynamically similar but have different energy balances, such as acceleration at high and low SOC, which differ in power consumption characteristics and the operation of the energy management system (BMS).
2. Identify sections where SOC stabilizes or increases, which may indirectly indicate recuperation activity or drive power reduction to save energy.
3. Detecting changes in the PHEV control strategy—during low SOC phases, the system can further reduce electric power and switch to fuel-efficient modes more frequently, even under similar speed and acceleration conditions.
3. Enhance the classification with energy consumption, allowing for analysis not only of vehicle behavior but also of the hybrid system’s efficiency depending on terrain conditions and driving style.

From a statistical analysis perspective, adding SOC does not significantly increase cluster separability (the silhouette increases minimally), as SOC changes slowly and has low variance in a short road test. However, from an engineering and energy interpretation perspective, its presence is very valuable—it provides each cluster with an informative dimension related to the actual energy flow within the vehicle.

Introducing a fifth parameter does not radically improve the mathematical quality of clustering, but it significantly increases its physical and diagnostic utility. This allows cluster analysis to more fully reflect the operation of a plug-in hybrid vehicle, combining kinematics, topography, and energy in a single descriptive model.

By increasing the number of state parameters subjected to clustering, the next logical step seems to be to increase the number of clusters. Based on the silhouette index, we decided to divide the clusters into seven clusters. The results of clustering in the 5-state space into seven clusters are presented in Figure 13.

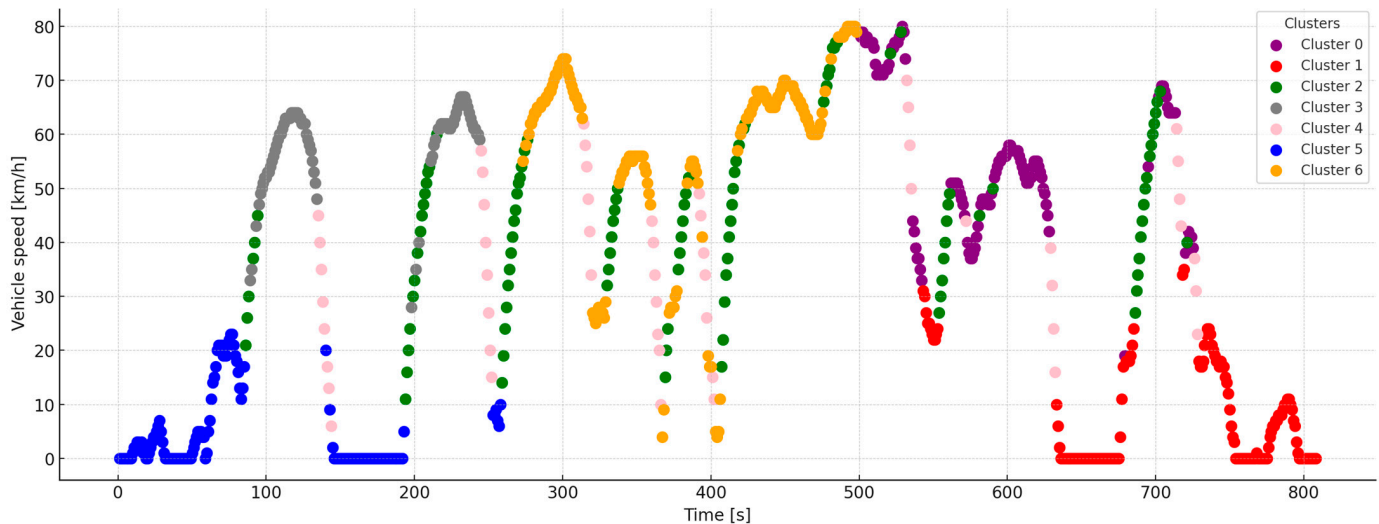


Figure 13. Results of clustering in a 5-state space divided into 7 states (clusters) using the K-Means algorithm.

Increasing the number of clusters to seven ($k = 7$) in the analysis based on five state parameters—electric motor power, speed, acceleration, altitude above sea level, and battery charge level (SOC)—allows for a more detailed, multidimensional picture of vehicle behavior during driving (see Figure 14). Compared to the five-cluster model, clustering into seven groups allows for the detection of subtle differences in the drivetrain's operating characteristics and a more accurate representation of energy-kinematic changes in diverse terrain.

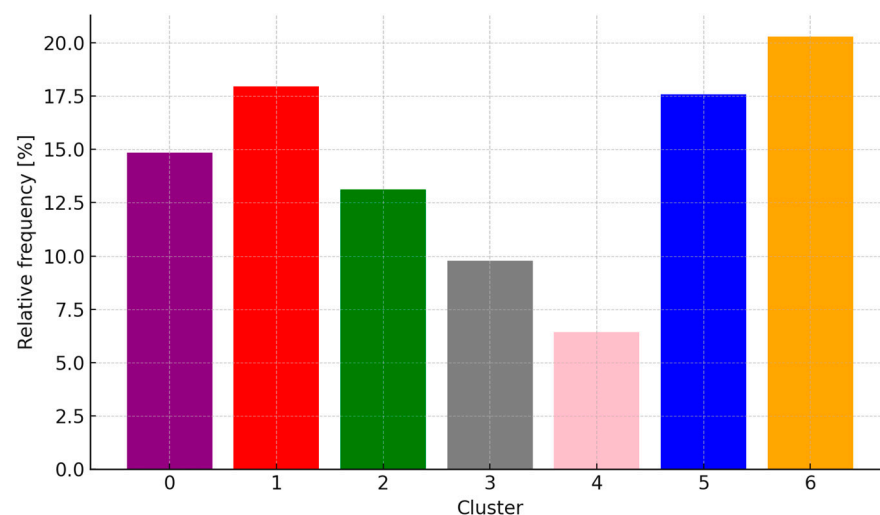


Figure 14. Relative frequencies of occurrence of individual 7 clusters.

Here is what additional information can be obtained:

1. Separation of acceleration and braking intensity—previously, a single cluster could encompass both light and heavy acceleration; now, for example, a standstill start (low power, positive acceleration) can be distinguished from dynamic acceleration (high

power, high acceleration). Similarly, gentle deceleration can be distinguished from deep regenerative braking.

2. Detecting transient states—with a larger number of clusters, groups appear corresponding to transition periods between driving phases, such as between acceleration and a steady-state phase. Such clusters are valuable for assessing driving smoothness and electric drive control strategies.
3. Identifying the influence of topography—thanks to the altitude parameter, it is possible to separate driving on hills from driving on flat terrain, even at similar speeds and power levels [29]. This allows for analyzing how changes in terrain affect energy consumption and recuperation frequency.
4. Distinguishing between SOC-dependent states—the 7-cluster model often distinguishes clusters associated with different battery charge levels, allowing for the study of how the control strategy changes with high vs. low SOC (e.g., power reduction, increased frequency of gentle energy recovery phases).
5. Better representation of real-world driving cycles—more clusters enable the creation of a more detailed “driving pattern” that can be used to compare different vehicles, drivers, or routes.
6. Enabling the detection of anomalies or unusual drivetrain behavior, such as situations where the vehicle consumes too much energy at low SOC or reaches high power on descents—such outliers can reveal control inefficiencies or component degradation.
7. Increased resolution of the energy-dynamic space—clustering with a larger k enables the study of detailed drivetrain operating regimes: from economical driving, through comfortable driving, to dynamic driving, which is important for calibrating driving strategies in future PHEV models.

The move from 5 to 7 clusters not only increases the resolution of the analysis but also allows for a better understanding of the complex relationships between driving dynamics, topography, and the vehicle’s energy state. From the point of view of transport research and energy efficiency of PHEVs, this level of detail can be crucial for building predictive models of energy consumption, analyzing driver driving style, and optimizing energy management in hilly terrain.

Below is a description of the relative frequency of clusters ($k = 7$) along with an interpretation of their operational significance:

1. Cluster 0 (purple)—14.9%: A typical section of a smooth start from a traffic light, driving in a traffic jam, or slowly rolling through urban zones at 30 km/h.
2. Cluster 1 (red)—18.0%: Stopping at a traffic light, in traffic, waiting to turn. Typical “idle travel” in EV mode, where the vehicle waits without increasing power consumption.
3. Cluster 2 (green)—13.2%: Corresponds to smooth, steady driving on longer straights, e.g., exit streets, sections between intersections, and city roads at a speed of 50 km/h.
4. Cluster 3 (gray)—9.8%: A typical situation when approaching a traffic light, pedestrian crossing, or the end of a line of vehicles. Strongly associated with energy recovery and power recuperation.
5. Cluster 4 (pink)—6.4%: Responds to emergency braking situations, e.g., when the driver reacts to changing lights, pedestrian traffic, merging drivers or the sudden stop of a column of cars.
6. Cluster 5 (blue)—17.6%: This is dynamic starting—for example, after changing traffic lights, joining traffic, leaving a minor lane, or wanting to join traffic before other vehicles.

7. Cluster 6 (orange)—20.3%: This involves driving on sections of high-speed traffic, such as city bypasses, dual carriageways in cities, and sections at 70 or 80 km/h. In real-world EV-PHEV data, such sections often appear between districts.

Overall, the picture is consistent for urban driving in hilly terrain: phases of smooth driving and acceleration dominate (clusters 0, 2, and 6), shorter braking episodes (clusters 3 and 4), and sections of slow movement and low-speed maneuvers (cluster 1), with additional moderately dynamic transients (clusters 4 and 5).

4. Discussion

Due to the complexity of the topic and the relatively large scope of research and analysis performed, the discussion of the obtained results will be presented in several areas. All interpretations presented in this section refer exclusively to the single recorded drive cycle. The findings should therefore be understood as illustrative examples of what the clustering method can reveal, rather than generalized behavioral signatures of PHEV operation.

4.1. Interpreting Clusters in the Context of Energy Efficiency of Electric Driving

Analysis of the clusters obtained in the study clearly indicates that the energy efficiency of PHEV electric driving is directly dependent on the pattern of occurrence of individual operating states. It was shown that the largest share of the total test time is occupied by clusters corresponding to smooth, steady driving (green and blue) and moderate acceleration (pink and gold), which together account for over half of the driving time. These sections are characterized by low energy consumption and a stable SOC level, confirming their dominant impact on the vehicle's energy balance. In contrast, clusters associated with dynamic acceleration and intensive braking occur less frequently but generate the largest power and acceleration fluctuations. While these clusters are smaller in proportion, their presence in the driving cycle indicates peak energy consumption and occasional recovery during deceleration.

The relative frequency of cluster occurrence indicates that the vehicle's driving pattern in urban conditions is moderately dynamic, which helps maintain the electric drive within the high-efficiency range. These results confirm that long-term operation in clusters corresponding to smooth driving is crucial for minimizing energy losses and extending the EV range. At the same time, short braking episodes, although constituting only a few percent of the driving time, play an important compensatory role—allowing for partial recovery of kinetic energy and stabilizing the battery charge level. In the case studied, this driving structure indicates a good correlation between the characteristics of urban traffic in Lublin and the capabilities of the plug-in hybrid drive, particularly in maintaining energy balance without excessive losses.

Ultimately, the clustering results demonstrate that the greatest energy efficiency potential of a PHEV is revealed in clusters corresponding to smooth, steady driving, while dynamic states, although shorter, determine the instantaneous load and power demand. This structure allows for the determination of the vehicle's actual usage profile and provides a starting point for further assessment of its efficiency in urban traffic.

4.2. Application of Clustering to Detect Energy Recovery

The clustering results show that regenerative braking phases can be inferred indirectly despite the absence of negative power values in the OBD data, which is consistent with previous findings on PHEV recuperation monitoring. Deceleration-dominated clusters, characterized by negative acceleration and near-zero electric motor power, align with the expected signature of energy recovery described in the literature. Short, recurring braking

episodes observed in the time series correspond to typical urban recuperation patterns, where kinetic energy is partially restored at medium vehicle speeds. The inclusion of altitude additionally supports the identification of downhill-related deceleration, confirming the terrain-dependent nature of recuperation emphasized in earlier studies. Overall, the clustering method provides a reliable qualitative indicator of energy recovery events, demonstrating that indirect kinematic cues can effectively capture recuperative behavior in PHEVs, as also noted by other researchers.

4.3. The Influence of Topography (Altitude Above Sea Level) on the Cluster Structure

Including altitude in the analysis enriched the clustering results by distinguishing whether observed speed and acceleration changes were driver-induced or terrain-induced, which aligns with previous findings on the topographic sensitivity of hybrid electric drive-trains. Uphill segments were associated with clusters showing higher power demand, while downhill sections corresponded to deceleration-dominated clusters, indicating potential recuperation episodes consistent with terrain-driven energy flow patterns described in the literature. This additional topographic dimension improved the physical interpretability of clusters, revealing how road gradients influence both load variability and the frequency of energy recovery, supporting earlier observations on real-world PHEV efficiency in hilly environments.

4.4. Comparison of Models with Different Numbers of Clusters

The comparison of models with different numbers of clusters shows that increasing k only marginally improves mathematical separation, as indicated by silhouette trends, while substantially increasing interpretive complexity. Although six- and seven-cluster models capture additional transitional driving states, many of these groups have low frequency or limited physical meaning, consistent with earlier observations on diminishing returns in high-resolution clustering. Overall, the five-cluster model offers the best balance between separability and interpretability, providing a stable representation of the main operating states without unnecessary fragmentation, in line with methodological recommendations reported in the literature.

4.5. Valorization of Research Methodology—OBD + KNIME

The adopted methodology, combining OBD-based data acquisition with processing in KNIME, proved effective for characterizing real-world operating states of the PHEV, offering a low-cost and accessible framework for vehicle energy analysis. Despite limitations of the OBD interface—such as restricted access to high-voltage system parameters and the 1 Hz sampling that may omit rapid transients—the workflow ensured consistent preprocessing, reproducibility, and clear interpretability of the resulting clusters. Overall, the OBD + KNIME approach represents a practical compromise between measurement simplicity and analytical capability, enabling reliable extraction of driving-state signatures in real urban conditions, as noted in previous methodological evaluations [30].

4.6. Practical Application of the Results

The identified clusters form a practical foundation for deriving operational signatures that characterize the energy and dynamic behavior of a PHEV during real-world urban driving. Such signatures can be used to evaluate driving efficiency, compare routes, or support eco-driving feedback by identifying patterns associated with high power demand or frequent braking. In fleet-management contexts, differences in cluster distributions may indicate driver-specific or route-specific inefficiencies, enabling targeted interventions or optimization of vehicle deployment. Moreover, the cluster-based representation can support simplified predictive models of energy consumption by linking average power

levels to the relative duration of operating states. These findings also demonstrate potential applicability in intelligent mobility systems, where real-time cluster detection could inform adaptive control strategies aimed at reducing energy use and enhancing the effectiveness of recuperation.

4.7. The Importance of the Obtained Research Results for Sustainable Transport

The results obtained in the conducted research indicate that plug-in hybrid vehicles have significant potential in building a sustainable transport system, particularly in urban traffic. The cluster structure revealed that the majority of the vehicle's driving time is spent in moderate load and smooth-running phases, which helps reduce energy consumption and minimize emissions. Maintaining electric drive in this range confirms the possibility of a significant portion of journeys in zero-emission mode, especially in urban environments with moderate traffic. The recorded power and acceleration values suggest that the vehicle operates in a high-efficiency regime, and the share of intense acceleration and braking phases is relatively small. This means that driving in electric mode is not only technically feasible but also effective in everyday use.

From the perspective of sustainable mobility, it is also significant that the vehicle was charged from a photovoltaic carport, further reducing the carbon footprint of the entire lifecycle. The research results confirm that, under favorable climatic and topographical conditions, it is possible to achieve a significant share of electric driving in the total mileage of a PHEV. Cluster analysis also shows that the hybrid drive system handles altitude changes well, maintaining stable energy consumption even in hilly terrain. This result confirms that plug-in hybrid technology can effectively support the transition to low-emission road transport, serving as an intermediate step between combustion and fully electric vehicles.

The limitations of the method include the fact that data from the OBD interface does not include the actual electricity balance. The lack of direct measurement of recuperative power and battery energy flows limits the ability to accurately assess environmental efficiency. Despite this, cluster analysis provided reliable information on vehicle traffic patterns and dynamics, allowing for a qualitative assessment of its environmental impact. The obtained results indicate that, with appropriate energy management and appropriate driving style, plug-in hybrid vehicles can be a viable tool for reducing emissions in cities while maintaining high user comfort and energy availability.

4.8. Additional Analytical Insights

The deeper inspection of cluster behavior highlights how driving dynamics, topography, and the vehicle's control strategy jointly shape the energy profile of the PHEV, confirming trends reported in prior studies on hybrid electric drivetrains. The dominance of steady-state clusters reflects the system's tendency to operate in a high-efficiency region, while high-power acceleration clusters reveal the characteristic rapid-torque response of electric motors described in the literature. The interaction between altitude and acceleration further demonstrates that road gradients significantly influence instantaneous energy demand and recuperation potential, emphasizing the importance of analyzing real-world rather than laboratory conditions. Collectively, these insights show that the identified clusters may also serve as a basis for simplified predictive models of energy consumption, linking cluster frequency with power distributions as suggested in earlier methodological work.

4.9. Fundamental Limitations

A fundamental limitation of this study is the very narrow scope of the dataset. The analysis was based on a single driver, a single vehicle model, a single predefined route,

and a relatively short driving duration. As a result, the identified clusters cannot be interpreted as universally representative “operational signatures” of PHEVs in electric mode. Instead, they should be viewed as route-specific and driver-specific behavioral patterns that illustrate the feasibility of applying unsupervised clustering to OBD-derived time series rather than providing generalizable quantitative benchmarks.

The restricted dataset also limits the ability to capture inter-driver variability, inter-vehicle differences in control strategies, the effect of weather and traffic, and long-term SOC evolution—all of which may significantly affect cluster composition and frequency. Furthermore, the short test duration reduces the likelihood of observing rare operating states or statistically stable distributions. For this reason, the results do not allow for broader conclusions about energy efficiency implications but only demonstrate how such implications could be inferred if richer datasets were available.

Therefore, the findings presented in this work should be treated as proof of concept, showing that clustering-based signatures are technically achievable using low-cost OBD telemetry but not yet validated against a sufficiently large or diverse sample. Future research must expand the dataset to include multiple drivers, repeated trials, varied traffic conditions, different PHEV models, and longer driving periods to ensure statistical robustness and enable generalization to real-world mobility studies.

It should be noted that the objective of this study is methodological rather than generalizable: the operational states identified in this work describe only the analyzed trip and are not intended to represent universal driving patterns for PHEVs.

5. Conclusions

This study presented a methodological framework for the analysis of energy-related operating states of a plug-in hybrid electric vehicle based on limited and incomplete on-board diagnostic (OBD) data. The framework was illustrated using a single short real-world urban drive, deliberately treated as a constrained and non-representative realization of a transportation process rather than as a basis for statistical generalization.

The results demonstrate that, even under severely limited observability and in the absence of direct measurements of key variables such as regenerative braking power, it is possible to recover physically interpretable operating states through an appropriate decomposition of the state space and unsupervised clustering. By progressively expanding the dimensionality of the analyzed feature space and evaluating multiple clustering configurations, the proposed approach enables the identification and interpretation of internally consistent energy-related states within the analyzed drive.

Importantly, the presented findings are strictly limited to the specific driving realization considered in this study. No claims are made regarding the typicality of the observed operating states, their frequency of occurrence, or their applicability to other vehicles, drivers, routes, or traffic conditions. The study does not aim to establish population-level benchmarks or generalized efficiency metrics but rather to demonstrate a principled approach to information recovery from constrained real-world OBD measurements.

The methodological contribution of this work lies in showing that a single, short, and incomplete OBD time series can still contain sufficient internal structure to support meaningful physical interpretation when analyzed using a carefully designed clustering-based framework. As such, the proposed approach may serve as a useful analytical tool for exploratory studies of vehicle operation in real-world conditions where comprehensive telemetry is unavailable and where classical data-rich methods cannot be readily applied.

Future work will focus on applying the proposed framework to a broader set of real-world driving scenarios, including multiple trips, routes, and vehicles, in order to assess the robustness and transferability of the identified operating-state structures. Such exten-

sions will enable the evaluation of how the methodological findings presented here scale under increasing data availability, without altering the fundamental inference principles demonstrated in this study.

Author Contributions: Conceptualization, M.L.; methodology, B.M.-S.; software, K.P.; validation, A.M. and B.Š.; formal analysis, J.C.; investigation, A.M.; resources, J.C.; data curation, B.Š.; writing—original draft preparation, M.L. and A.M.; writing—review and editing, A.M.; visualization, J.C.; supervision, A.M.; project administration, A.M. and K.P.; funding acquisition, B.Š. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V05-00002.

Data Availability Statement: The original contributions presented in this study are included in the article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BEV	Battery Electric Vehicle
BMS	Battery Management System
CSV	Comma-Separated Values
ECU	Electronic Control Unit
EV	Electric Vehicle
GDI	Gasoline Direct Injection
ICE	Internal Combustion Engine
KNIME	Konstanz Information Miner (Analytics Platform)
OBD	On-Board Diagnostics
PHEV	Plug-in Hybrid Electric Vehicle
PV	Photovoltaic
SOC	State of Charge
TLA	Three-Letter Acronym
WLTC	Worldwide Harmonized Light Vehicles Test Cycle

References

1. Michailidis, E.T.; Panagiotopoulou, A.; Papadakis, A. A Review of OBD-II-Based Machine Learning Applications for Sustainable, Efficient, Secure, and Safe Vehicle Driving. *Sensors* **2025**, *25*, 4057. [\[CrossRef\]](#)
2. Ramai, C.; Ramnarine, V.; Ramharack, S.; Bahadoorsingh, S.; Sharma, C. Framework for Building Low-Cost OBD-II Data-Logging Systems for Battery Electric Vehicles. *Vehicles* **2022**, *4*, 1209–1222. [\[CrossRef\]](#)
3. Szumska, E.M. Regenerative Braking Systems in Electric Vehicles: A Comprehensive Review of Design, Control Strategies, and Efficiency Challenges. *Energies* **2025**, *18*, 2422. [\[CrossRef\]](#)
4. Rykała, M.; Grzelak, M.; Rykała, Ł.; Voicu, D.; Stoica, R.-M. Modeling Vehicle Fuel Consumption Using a Low-Cost OBD-II Interface. *Energies* **2023**, *16*, 7266. [\[CrossRef\]](#)
5. Małek, A.; Caban, J.; Dudziak, A.; Marciniak, A.; Vrábel, J. The Concept of Determining Route Signatures in Urban and Extra-Urban Driving Conditions Using Artificial Intelligence Methods. *Machines* **2023**, *11*, 575. [\[CrossRef\]](#)
6. Małek, A.; Marciniak, A.; Kroczyński, D. Defining Signatures for Intelligent Vehicles with Different Types of Powertrains. *World Electr. Veh. J.* **2025**, *16*, 135. [\[CrossRef\]](#)
7. Hamza, K.; Laberteaux, K.P. Utility Factor Curves for Plug-in Hybrid Electric Vehicles: Beyond the Standard Assumptions. *World Electr. Veh. J.* **2023**, *14*, 301. [\[CrossRef\]](#)
8. Pielecha, I.; Cieslik, W.; Sz wajca, F. Energy Flow and Electric Drive Mode Efficiency Evaluation of Different Generations of Hybrid Vehicles under Diversified Urban Traffic Conditions. *Energies* **2023**, *16*, 794. [\[CrossRef\]](#)
9. Nazari, M.; Hussain, A.; Musilek, P. Applications of Clustering Methods for Different Aspects of Electric Vehicles. *Electronics* **2023**, *12*, 790. [\[CrossRef\]](#)

10. Cao, B.; Xing, Q.; Yang, K.; Wu, X.; Li, L. Unsupervised Contrastive Learning for Time Series Data Clustering. *Electronics* **2025**, *14*, 1660. [\[CrossRef\]](#)
11. Szumska, E.M.; Jurecki, R. The Analysis of Energy Recovered during the Braking of an Electric Vehicle in Different Driving Conditions. *Energies* **2022**, *15*, 9369. [\[CrossRef\]](#)
12. Kozłowski, E.; Zimakowska-Laskowska, M.; Dudziak, A.; Wiśniowski, P.; Laskowski, P.; Stankiewicz, M.; Śnauko, B.; Lech, N.; Gis, M.; Matijošius, J. Analysis of Instantaneous Energy Consumption and Recuperation Based on Measurements from SORT Runs. *Appl. Sci.* **2025**, *15*, 1681. [\[CrossRef\]](#)
13. Kropiwnicki, J.; Gawłás, T. Estimation of the Regenerative Braking Process Efficiency in Electric Vehicles. *Acta Mech. Autom.* **2023**, *17*, 303–310. [\[CrossRef\]](#)
14. Gou, Y. Research on Electric Vehicle Regenerative Braking System and Energy Recovery. *Int. J. Hybrid Inf. Technol.* **2016**, *9*, 81–90. [\[CrossRef\]](#)
15. Cai, W.; Liu, C. Long Downhill Braking and Energy Recovery of Pure Electric Commercial Vehicles. *World Electr. Veh. J.* **2024**, *15*, 51. [\[CrossRef\]](#)
16. Enang, W.; Bannister, C. Modelling and Control of Hybrid Electric Vehicles: A Comprehensive Review. *Renew. Sustain. Energy Rev.* **2017**, *74*, 1210–1239. [\[CrossRef\]](#)
17. León, R.; Montaleza, C.; Maldonado, J.L.; Tostado-Véliz, M.; Jurado, F. Hybrid Electric Vehicles: A Review of Existing Configurations and Thermodynamic Cycles. *Thermo* **2021**, *1*, 134–150. [\[CrossRef\]](#)
18. Junthopas, W.; Wongoutong, C. Pre-Determining the Optimal Number of Clusters for k-Means Clustering Using the Parameters Package in R and Distance Metrics. *Appl. Sci.* **2025**, *15*, 11372. [\[CrossRef\]](#)
19. Kozłowski, E.; Wiśniowski, P.; Gis, M.; Zimakowska-Laskowska, M.; Borucka, A. Vehicle Acceleration and Speed as Factors Determining Energy Consumption in Electric Vehicles. *Energies* **2024**, *17*, 4051. [\[CrossRef\]](#)
20. Chaudhry, M.; Shafi, I.; Mahnoor, M.; Vargas, D.L.R.; Thompson, E.B.; Ashraf, I. A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective. *Symmetry* **2023**, *15*, 1679. [\[CrossRef\]](#)
21. Akogul, S.; Erisoglu, M. An Approach for Determining the Number of Clusters in a Model-Based Cluster Analysis. *Entropy* **2017**, *19*, 452. [\[CrossRef\]](#)
22. He, Z.; Jia, Z.; Zhang, X. A Fast Method for Estimating the Number of Clusters Based on Score and the Minimum Distance of the Center Point. *Information* **2020**, *11*, 16. [\[CrossRef\]](#)
23. Yan, B.; Yin, Y.; Liu, P. A New Cluster Validity Index Based on Local Density of Data Points. *Axioms* **2025**, *14*, 578. [\[CrossRef\]](#)
24. Skuza, A.; Jurecki, R.; Szumska, E. Influence of Traffic Conditions on the Energy Consumption of an Electric Vehicle. *Commun. —Sci. Lett. Univ. Zilina* **2023**, *25*, B22–B33. [\[CrossRef\]](#)
25. Gechev, T.; Mruzek, M.; Barta, D. Comparison of Real Driving Cycles and Consumed Braking Power in Suburban Slovakian Driving. *MATEC Web Conf.* **2017**, *133*, 02003. [\[CrossRef\]](#)
26. Rizzo, G.; Naghinajad, S.; Tiano, F.A.; Marino, M. A Survey on Through-the-Road Hybrid Electric Vehicles. *Electronics* **2020**, *9*, 879. [\[CrossRef\]](#)
27. Weiss, M.; Winbush, T.; Newman, A.; Helmers, E. Energy Consumption of Electric Vehicles in Europe. *Sustainability* **2024**, *16*, 7529. [\[CrossRef\]](#)
28. Tarout, H.; Zaki, H.; Chahbouni, A.; Ennajih, E.; Louragli, E.M. Optimizing Energy Consumption in Electric Vehicles: A Systematic and Bibliometric Review of Recent Advances. *World Electr. Veh. J.* **2025**, *16*, 577. [\[CrossRef\]](#)
29. Dávila-Sacoto, M.; Toledo, M.A.; Hernández-Callejo, L.; González, L.G.; Alvarez Bel, C.; Zorita-Lamadrid, Á.L. Location of Electric Vehicle Charging Stations in Inter-Andean Corridors Considering Road Altitude and Nearby Infrastructure. *Sustainability* **2023**, *15*, 16582. [\[CrossRef\]](#)
30. KNIME Analytics Platform. Available online: <https://www.knime.com/> (accessed on 12 December 2025).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.