National Technical University of Athens
School of Civil Engineering
Department of Transportation Planning and Engineering

A Multilevel Integrated Assessment of Safe and Green Mobility

Virginia Petraki

Doctoral Dissertation

Supervising Committee

George Yannis, Professor NTUA - Supervisor
Eleni Vlahogianni, Professor NTUA
Athanasios Theofilatos, Assistant Professor UTH

Extended Synopsis

Objectives and Methodology

Sustainable mobility is a multidimensional phenomenon encompassing road safety, as well as economic and environmental performance. Existing methodologies have attempted to address these mobility dimensions individually, despite the potential synergies and trade-offs that might be uncovered when addressing them jointly. This lack of co-consideration underscores the need for analytical frameworks that holistically capture those dimensions. The key to this integrated approach is driving behavior, constituting the most critical factor that affects both safe and green mobility.

Given the above context, the main objective of this dissertation is the **multilevel integrated assessment of safe and green mobility**. This assessment is conducted across multiple levels, from individual trips to the road network, by fusing road infrastructure, traffic, and weather-related data with high-resolution naturalistic driving behavior data.

Initially, a detailed systematic review was carried out aiming to explore the joint effects of driving behavior on road safety, fuel consumption, and vehicle emissions in the context of an integrated sustainability approach. An innovative **two-stage PRISMA** methodology was developed, identifying 78 commonly used behavioral and contextual factors, data collection methods, and modeling techniques across safety and eco-driving studies, while also revealing a clear research gap in integrated safety-eco analytical frameworks. Thus, these findings highlight the importance of developing comprehensive analytical frameworks that holistically assess safe and green mobility.

Subsequently, the following **research questions** were formulated:

Research Question 1
How can driving behavior, road infrastructure, road crashes, traffic, and weather datasets be fused to enable the integrated assessment of safe and green mobility at trip and spatial levels? Which indicators are most appropriate to capture safe and green mobility across levels?

Research Question 2
How can sustainable trip patterns be identified by integrating Surrogate Safety Measures (SSMs) and fuel consumption? Is it possible to predict and explain them through behavioral and contextual features, and, if so, which Machine Learning (ML) classifiers are the most appropriate?

Research Question 3
How can crash risk and fuel consumption hotspot spatial patterns be systematically identified and spatially compared across road junctions? Which road junctions achieve the best trade-off between crash risk and fuel consumption?

Research Question 4
How can safe and green driving outcomes be jointly modeled at the trip and road segment levels? Do they share common mechanisms that explain their divergence, or co-occurrence across levels?

Research Question 5
How can sustainable driving efficiency be assessed at the trip level, by integrating safety, fuel consumption, and travel time, and translated into road efficiency? Which exogenous factors influence sustainable driving efficiency?

These research questions served as the conceptual and methodological backbone of this dissertation, structuring the development of an integrated assessment of safe and green mobility. To provide scientifically supported answers, an elaborate analytical methodological framework was developed, which is presented on Figure I.

The analytical framework of this dissertation comprises three complementary analytical modules with distinct but interrelated objectives.

Module 1

Pattern Identification of Safe and Green Mobility:

The first module identifies sustainable mobility patterns at the individual trip level and at the road junction level, by incorporating road safety and fuel consumption indicators. The identified trip and spatial patterns provide a foundation for understanding safe and green mobility interactions across levels, informing the subsequent modules.

Module 2

Joint Modeling of Safe and Green Driving Outcomes:

The second module advances towards the quantification of the common mechanisms influencing surrogate safety and fuel consumption, at both individual trip and road segment levels. Joint structural modeling of safe and green driving outcomes enables the direct and indirect effects of driving behavior, roadway, and environmental factors.

Module 3

Sustainable Driving Efficiency Assessment:

The third module reconceptualizes sustainable driving, unifying surrogate safety, fuel consumption, and travel time into a single efficiency score transferable from individual trips to the road network. The internal structure of these efficiency scores is further investigated, alongside the external contextual factors influencing efficiency.

The core of the methodological framework involved a multistage process, starting with the **investigation and collection of data exploited in this dissertation**, as well as the development of appropriate data fusion procedures for ensuring consistency in levels and dimensions within each analytical module. High-resolution (per second) naturalistic driving data were recorded within the Attica Region in Greece, which was defined as the research area, amounting to 35,637 trips, and providing vehicle kinematics data, including speed and acceleration profiles, and SSMs such as harsh acceleration and braking events, speeding (exceeding the speed limit), and mobile phone distraction. To complement the SSMs, 10,893 historical road crash records from 2016 to 2022, consistently attributed to 5,517 junctions within the study area. Following, the high-resolution telematics data were systematically map-matched to 34,889 road segments, through OpenStreetMap (OSM), allowing the simultaneous integration of driving behavior, road infrastructure characteristics including road grade. Instantaneous fuel consumption was estimated through the Virginia Tech Comprehensive Power-Based Fuel Consumption Model (VT-CPFM) model, using the map-matched high-resolution kinematics data. Traffic conditions were depicted through traffic speeds retrieved from Google for 42 critical roads and specific time periods in the study area. Next, weather conditions, such as temperature, relative humidity, precipitation, and daylight, were temporal matched with the telematics data.

Considering the multilevel character of the analytical framework, **four databases were constructed**, namely, a trip-level database (used in the 1st and 2nd modules), a road segment-level database (used in the 2nd module), a junction-level database (used in the 1st module), and a road-based trip database, which was restricted to trips where traffic information was available (used in the 3rd module). The creation of these databases involved the implementation of three different spatiotemporal data fusion procedures including (1) a map-matching procedure to match each trip-second to the corresponding road segment, (2) a second map-matching procedure to link historical crash records and high-resolution driving data with specific road junctions, and (3) a third spatiotemporal matching process between individual trip data and broader traffic data.

The following stage included the selection of **statistical and analytical tools** capable of capturing the multidimensional and multilevel nature of the integrated assessment of safe and green mobility in each analytical module. In the first analytical module, trips were sequentially profiled based on (i) road type exposure, (ii) SSMs, and (iii) fuel consumption, using unsupervised techniques such as k-means clustering. The predictability of the identified sustainable trip profiles was determined through the implementation of supervised ML techniques such as Random Forest (RF), Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), and eXtreme Gradient Boosting (XGBoost). The performance of the ML models was determined through cross-validation metrics, while model interpretability was enhanced using SHapley Additive exPlanations (SHAP). At the spatial level, spatial clustering and hotspot analysis were conducted using global spatial measures such as Global Moran's *I* to measure overall spatial autocorrelation, as well as local spatial analysis such as LISA and Getis-Ord Gi*, to detect crash risk and fuel consumption hotspots and coldspots at road junction level. Cross-type L-function was further employed to examine the spatial association between safety and fuel hotspots, while a Pareto-based dominance analysis was applied to unveil the road junctions presenting the best trade-offs.

In the second analytical module, the joint modeling of safe and green driving outcomes was conducted using two Structural Equation Models (SEMs) developed at the trip and road segment levels, taking into account the need for spatial correction. Finally, in the third analytical module, sustainable trip efficiency was quantified through Data Envelopment Analysis (DEA) integrating SSMs, fuel consumption, and driving time into composite efficiency scores. The internal structure of the resulting efficiency scores was further explored through an explainable ML stage, using XGBoost algorithm and SHAP values. Given that the resulting efficiency scores were bound between zero and one, a mixed-effect beta regression model was developed to identify the external contextual determinants influencing trip efficiency. Collectively, this integrated analytical architecture enabled the systematic fusion of heterogeneous datasets and supported a coherent multilevel assessment of safe and green mobility.

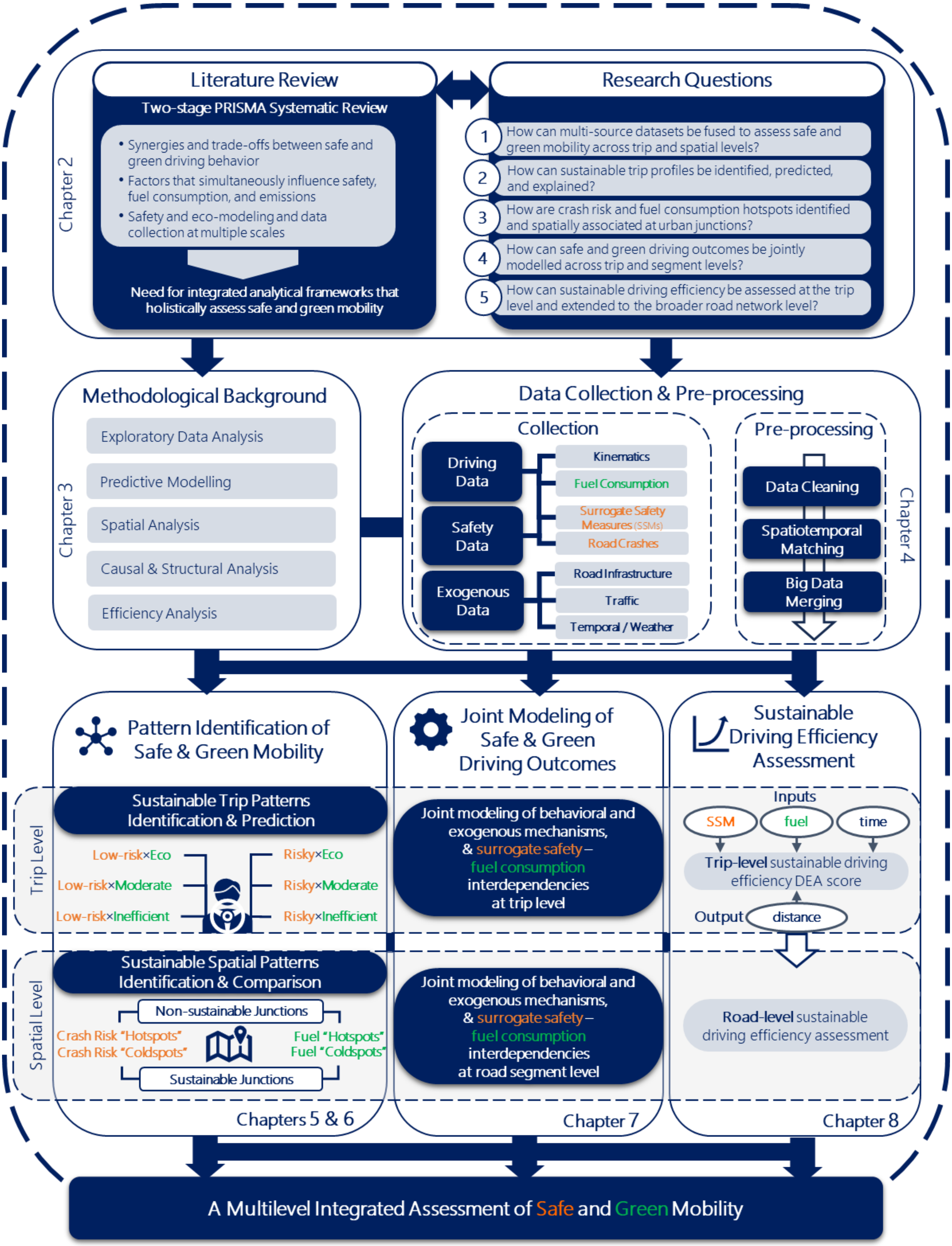

Figure I Overall methodological framework of the doctoral dissertation

Main Findings

Ultimately, the collective synthesis of all analyses carried out within this doctoral dissertation has led to an integrated assessment framework of safe and green mobility with a significant number of **original results**, as presented below.

Module 1 | Main Findings from Pattern Identification of Safe and Green Mobility

The first analytical module investigated how safe and green mobility patterns co-evolve across trips and spatial levels. The findings demonstrate that their relationship is neither uniformly synergistic nor inherently antagonistic; instead, their relationship is **mechanism-dependent and level-sensitive**.

Trip Level

At the trip level, **six sustainable trip profiles were identified**, namely:

1. “Low-risk×Eco”
2. “Risky×Eco”
3. “Low-risk×Moderate”
4. “Risky×Moderate”
5. “Low-risk×Inefficient”
6. “Risky×Inefficient”

The identified trip profiles aimed to capture distinct configurations of safety exposure and fuel consumption revealing structured synergies and trade-offs. Initially, it was revealed that aggressive and distracted trips consume up to 1.6 lt/100km more fuel than low-risk trips, while speeding in urban and rural contexts may reduce fuel consumption by up to 2.2 lt/100km despite increasing crash risk, while the opposite pattern tends to emerge for trips with higher motorway exposure, demonstrating that safety and fuel efficiency interactions are strongly context-dependent.

To predict the identified sustainable trip profiles, **four multinomial supervised ML classifiers were trained** and tested. XGBoost outperformed the other classifiers achieving 60% overall test accuracy, with “Low-risk×Eco” trip profile achieving the highest correct classification rate of 71.4%. SHAP results highlighted average speed, speed variability, stop frequency, route complexity including road grade, and junction exposure, as the most significant determinants, overshadowing the weather and temporal effects in shaping sustainable trip profiles.

Notably, **two trip profiles revealed considerable synergies**. The “Low-risk×Eco” profile represented the most sustainable driving behavior, combining low surrogate safety indicators and comparatively low fuel consumption (≈6 lt/100km), typically occurring under moderate speeds, low maximum speeds, and a low share of stopping time. In contrast, the “Risky×Inefficient” profile was characterized by the highest levels of mobile phone use and harsh events, combined with high fuel consumption, and was predominantly correlated with low average speeds, frequent stops, increased speed variability, and peak speeds, as well as driving on steeper road grades.

On the other hand, **two patterns of trade-off were identified**. The “Risky×Eco” profile achieved relatively low fuel consumption despite frequent speeding, characterized by high average and upper-quartile speeds, few stops, and limited motorway exposure, suggesting free-flow yet unsafe speed choices on urban roads. Conversely, the “Low-risk×Inefficient” profile showed the highest median fuel consumption (≈12.3 lt/100km) despite almost zero risky events, reflecting inefficiencies associated with low driving speed, frequent stops, limited upper-tail speeds, and driving on steeper road grades. The remaining profiles, "Low-risk×Moderate" and "Risky×Moderate", represented intermediate conditions with moderate fuel consumption and low risk driving.

Spatial Level

Regarding spatial patterns of safe and green mobility, a spatial analysis was conducted at 5,517 junctions integrating historical road crash data, and fuel consumption data. Crash risk (expressed as crashes per 1,000 trips) and fuel consumption (expressed as liters per 1,000 trips) both **exhibited statistically significant spatial clustering**, with Global Moran's *I* confirming positive spatial autocorrelation. Getis-Ord Gi* method identified, at the 99% confidence level, 182 crash risk and 232 fuel consumption hotspots within the study area. However, their direct overlap was minimal, suggesting that the most crash-prone junctions are not necessarily the ones where the most fuel-inefficient driving behavior occurs.

Spatial association between crash risk and fuel consumption hotspots was further examined using cross-type L-function showing a nonrandom spatial relationship. Results indicated the presence of short-range spatial segregation, meaning that the two hotspot patterns tend to occur within the same broader urban areas but also to segregate at short distances (<3km), occupying different nearby junctions. This association weakened and became negligible at larger distances (>6 km). Moreover, most crash risk hotspots are mainly located at junctions of tertiary roads, while fuel consumption hotspots seem to be more related to junctions where primary roads are involved.

Finally, **Pareto-based dominance analysis** complemented hotspot detection by identifying a set of non-dominated junctions that provide the best safety-fuel trade-offs, while a weighting-based sensitivity analysis was conducted to demonstrate how the ranking of these junctions shifts under safety-oriented versus efficiency-oriented decision scenarios. These junctions can serve as practical reference points for integrated strategies, since improvements at these locations are more likely to produce benefits for both objectives without creating major trade-offs.

Module 2 — Main Findings from Joint Modeling of Safe and Green Driving Outcomes

Two SEMs were developed at the trip and spatial levels, offering valuable insights into the influence of four latent variables related to driving volatility, roadway context, geometric characteristics, as well as exogenous weather and temporal conditions on driving surrogate safety and fuel efficiency. The SEMs displayed excellent goodness-of-fit statistics at the trip level (CFI = 0.947, TLI = 0.922, RMSEA = 0.060, SRMR = 0.037) and segment level (CFI = 0.947, TLI = 0.923, RMSEA = 0.044, SRMR = 0.036).

Trip Level

At the trip level, driving volatility, measured through stop frequency, stopping time share, and cruising time share, was the dominant determinant for both safe and green driving outcomes. Higher driving volatility was positively associated with fuel consumption, aggressive (increased harsh acceleration and braking events) and distracted driving (increased mobile phone use), and negatively associated with speeding. Greater exposure to primary roads was significantly associated with lower fuel consumption, aggressive and distracted driving but higher speeding frequency, compared to trips dominated by residential roads and intra-municipal trips. Also, flatter terrain and lower grade variability during the trip were significantly associated with reduced fuel consumption, aggressive and distracted driving but did not show a significant direct relationship with speeding once other factors were controlled. Favorable environmental conditions such as daytime operation, higher temperatures, and lower relative humidity slightly moderate unsafe and fuel inefficient driving behaviors. Residual covariances at the trip level showed strong positive associations between fuel consumption and aggressive driving and smaller but significant positive association with speeding and distracted driving. Overall, these findings indicate that, beyond contextual and operational factors, persistent trip-level behavioral influences contribute to the co-occurrence of fuel inefficiency and crash risk.

Spatial Level

At the road segment level, similar latent constructs were examined, yet the relative importance of mechanisms shifts. In this level of analysis, roadway characteristics followed by driving volatility exerted the strongest statistical influence. Driving volatility, measured through frequent stops and high speed variability among trips that traversed each road segment, remained positively associated with the aggregated fuel consumption, aggressive and distracted driving, and negatively associated with speeding. Primary road segments were characterized by lower fuel consumption and fewer harsh and mobile usage events, but higher speeding frequency compared with residential and one direction road segments. Steeper grades, shorter segment lengths, and higher intersection density substantially constrained speeding while increasing fuel demand, aggressive and distracted driving. Environmental and temporal conditions were less significant at the spatial level. The spatial correction analysis showed that spatial dependence existed in both latent constructs and driving outcomes, but accounting for it did not alter the statistically significant identified relationships. Residual associations differed from the trip level. Fuel consumption maintained a positive residual association with aggressive and distracted driving. However, fuel consumption and speeding displayed a small but statistically significant negative residual association, while aggressive driving was negatively correlated with speeding. This indicates that road segments characterized by elevated fuel consumption and frequent harsh acceleration and braking events are not the same segments where speed limit violations occur, reflecting a localized differentiation.

Overall, the results indicated that **fuel inefficiency aligns consistently with aggressive and distracted driving at both trip and spatial levels**, reflecting shared behavioral and contextual mechanisms. This association may be explained by the fact that harsh acceleration and braking events increase engine load and fuel demand, while distraction can destabilize longitudinal control. Speeding, however, is governed by partially opposite structural mechanisms compared to fuel consumption at both levels, yet at the trip level it still exhibits a positive residual covariance, indicating that they can co-occur probably due to shared, unmodeled trip level influences. In contrast, at the road segment level, local constraints prevent co-occurrence between fuel efficiency and low speeding.

These findings confirm that **safe and green outcomes are interconnected but not interchangeable**. Strategies aimed at reducing driving volatility and limiting exposure to geometrically complex road environments are likely to deliver co-benefits for both energy and road safety, while speeding requires a management that depends strongly on roadway type and geometry.

Module 3 — Main Findings from Sustainable Driving Efficiency Assessment

While the first module identified structured safe and green mobility patterns and the second quantified the mechanisms governing their interaction, this module **operationalizes sustainability into a measurable multi-criteria driving efficiency score**, bridging microscopic trip efficiency with macroscopic road efficiency. Initially, trip efficiency was assessed by developing an input-oriented Banker, Charnes, and Cooper (BCC) DEA, considering undesirable driving outcomes such as SSMs (e.g., harsh braking, mobile phone usage), fuel consumed (litres), and driving time (minutes) as inputs, and trip distance (kilometers) as the output.

The outcomes revealed that **efficient trips** were consistently characterized by lower fuel consumption, shorter driving time per kilometer, as well as fewer harsh braking events and mobile phone use. To further investigate the internal structure of the composite efficiency scores, XGBoost models combined with SHAP analysis were implemented and interpreted. Results revealed that driving time exerts the strongest influence on the efficiency score, followed by fuel consumption, while safety indicators operate primarily as penalty-type constraints, reflecting their relative sparsity. However, an additional DEA was conducted as a sensitivity analysis, restricted to trips exhibiting safety-related events, revealing a structural shift in explanatory contributions with SSMs gaining importance. Therefore, these findings highlight road safety indicators as more critical determinants of the efficiency score when exposure to crash risk is present.

Trip-level efficiencies were aggregated to derive a **road-level efficiency metric** for those roads with available traffic data. This aggregation allowed for the derivation of road network sustainable driving efficiencies, demonstrating the utility of such high-resolution telematics data for driving network-level assessment. The third module also explored the contextual factors developing a mixed-effects beta regression model, using as a response variable the trip efficiency score while accounting for route-level heterogeneity. Traffic congestion emerged as the most influential external determinant, compared with road infrastructure and temporal conditions, systematically reducing trip sustainability by increasing fuel use, driving time, as well as harsh braking and mobile phone use. Temporal and infrastructural variables also exhibited systematic effects. Overall, this analytical module demonstrated that sustainable driving efficiency is not an abstract concept but rather a quantified and interpretable performance indicator.

Innovative Scientific Contributions

This doctoral dissertation provides a series of major contributions to the field of sustainable mobility through the development of an integrated multilevel analytical framework for the assessment of safe and green mobility. The main scientific contributions of the dissertation are summarized in Figure II.

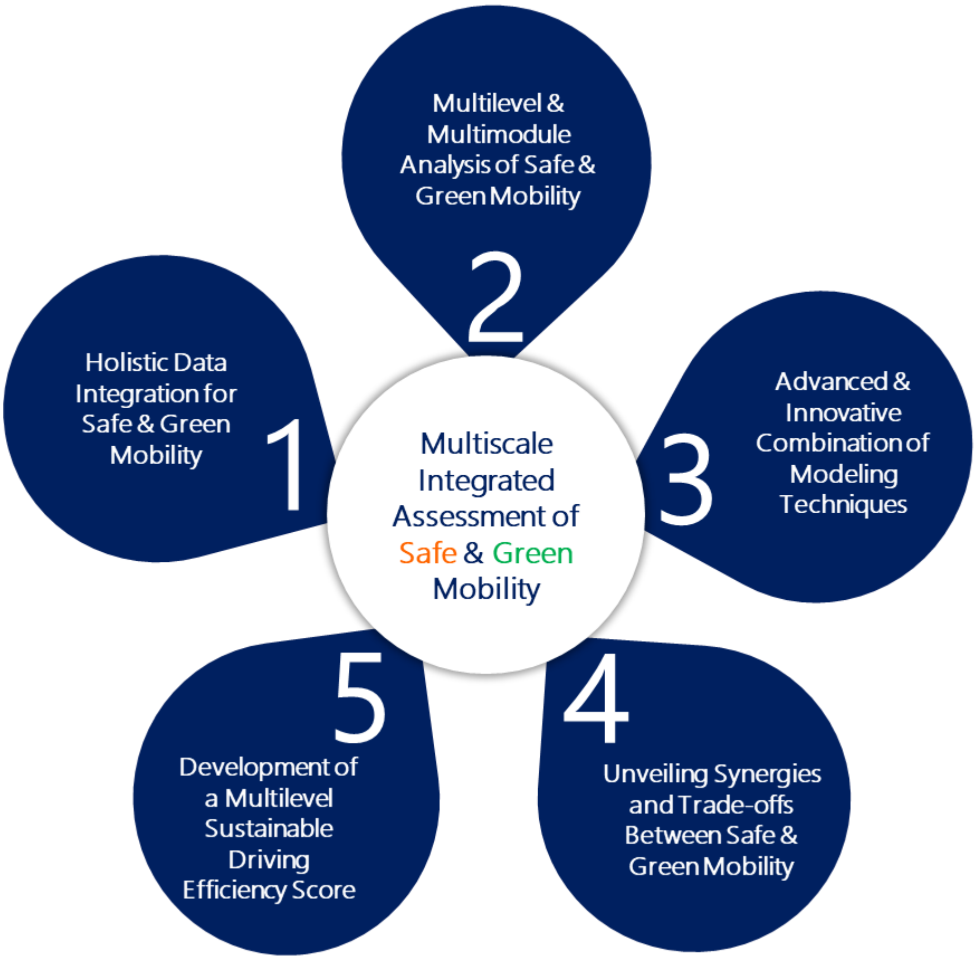

Figure II Innovative contributions of this doctoral dissertation

- ### Holistic Data Integration for Safe and Green Mobility

A key innovative contribution is the development of a **unified and multilevel data ecosystem** that operationalizes safe and green mobility within a single analytical framework. The innovation lies not in the individual datasets themselves, but in their structured fusion into a coherent, spatially and temporally consistent architecture. High-resolution telematics data were systematically map-matched through OSM, allowing the simultaneous integration of driving behavior, road infrastructure characteristics including road grade. Instantaneous fuel consumption was estimated through the VT-CPFM model using the map-matched kinematics data, enabling the direct alignment of SSMs and fuel consumption data at common temporal and spatial levels, which is rarely achieved in existing literature. This dissertation extends the telematics-based framework by embedding historical police-recorded crash data into the same spatial reference system. Given that crash location accuracy is highest at junction level, crashes, aggregated driving behavior indicators, and fuel consumption estimates were map-matched and merged at this spatial level, resulting in a novel dataset that links recorded crashes with driving and fuel metrics. Traffic and weather dimensions were not treated as peripheral controls but were structurally embedded into the data architecture. Traffic conditions were spatiotemporally aligned with telematics trajectories along selected critical roads, enabling the explicit examination of congestion impacts on joint safe and green driving outcomes. Similarly, temporal matching of weather data allowed the framework to capture variability induced by meteorological and temporal conditions.

- **Multilevel and Multimodule Analysis of Safe and Green Mobility**

A further innovative contribution is the development of a structured multimodule analytical framework that systematically investigates safe and green mobility across two complementary levels, namely the trip and the spatial level. Specifically, this framework is structured around three analytical modules, which combine **pattern identification, structural modeling of relationships, and multi-criteria efficiency assessment** within a unified analytical structure covering individual trip, road segment and junction levels. Thus, the interlinking of the developed modules within the same analytical framework enables an integrated assessment of safe and green mobility. Also, this vertical integration from individual trips to system-level performance represents a methodological advancement in sustainable mobility domain that remains relatively limited in existing literature.

- **Advanced and Innovative Combination of Modeling Techniques**

Another significant contribution is the advanced and **integrated combination of statistical, spatial, econometric, benchmarking and ML techniques** within a unified analytical framework. First, unsupervised clustering was combined with supervised classification to identify and predict combined trip profiles encompassing safety and eco dimensions. The coupling of trips clustering with explainable ML strengthens both predictive performance and transparency. Second, the SEM model at road segment level, further investigated through spatial correction procedures to account for spatial autocorrelation, representing a methodological advancement. Third, DEA was combined with explainable ML algorithms to create an interpretable multi-criteria driving efficiency score. Rather than using conventional DEA results, this dissertation attempted to integrate the use of SHAP interpretation and contextual analysis, using beta regression modeling, to enhance the explanatory power of non-parametric benchmarking modeling. Thus, this innovative contribution lies not simply in the individual application of these techniques, but in their combination, advancing the analytical sophistication of sustainable mobility research and providing a transferable modeling blueprint.

- **Unveiling Synergies and Trade-offs Between Safe and Green Mobility**

A major innovative contribution is related to the **data-driven integrated assessment** of road safety and economic-environmental performance through fuel consumption, addressing them as interrelated dimensions of sustainable mobility.

This dissertation provides statistically supported evidence that interactions related to road safety and fuel efficiency are mechanism- and level-dependent, revealing synergies and context-sensitive trade-offs. At the trip level, it introduces combined sustainable mobility profiles that systematically capture synergistic (“Low-risk×Eco”) and critical (“Risky×Inefficient”) driving patterns. Using explainable ML techniques, it identifies dominant behavioral and contextual determinants including speed selection, stopping behavior, road environment complexity, thereby establishing an interpretable basis for driver feedback and incentive systems. At the spatial level, it developed an integrated hotspot and Pareto-based analytical framework for the simultaneous evaluation of crash risk and fuel consumption at the level of the road junction. It determines the spatial distribution of crash risk and fuel consumption hotspots, the areas of overlap between the two types of hotspots, and the road junctions with the best trade-offs between safety and efficiency, providing robust decision-support tools for planners and road authorities. Moreover, this dissertation offers insights into the mechanisms connecting surrogate safety and fuel efficiency. SEM is used to formally test and confirm these relationships, offering a data-driven understanding of when safety and efficiency goals reinforce each other and when trade-offs arise.

- **Development of a Multilevel Sustainable Driving Efficiency Score**

Another significant innovative contribution is the estimation of a **unified sustainable driving efficiency score**, which coherently integrates road safety, fuel, and driving time efficiency within a unified analytical framework, allowing for the comprehensive assessment of driving sustainability. By effectively combining multi-criteria efficiency modeling with explainable ML techniques, the developed approach enhances the interpretability of efficiency outcomes and enables the identification of the most influential determinants of sustainable driving performance. Another major contribution is in vertical scalability from microscopic trip-level efficiency scores to the macroscopic road-level efficiency rankings, linking individual driving behavior and network-wide sustainability outcomes. From a practical perspective, the developed sustainable driving efficiency score can be utilized as a tool for monitoring and evaluating sustainable mobility, considering simultaneously road safety, the environment, and traffic.