CHALMERS
UNIVERSITY OF TECHNOLOGY

# Finding representative routes of a vehicle's operations

Master's thesis in engineering mathematics and computational science

Oliver Palada

DEPARTMENT OF MECHANICS AND MARITIME SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis in engineering mathematics and computational science

# Finding representative routes of a vehicle's operations

Oliver Palada

Department of Mechanics and Maritime Sciences
Division of Vehicle Engineering and Autonomous Systems
Chalmers University of Technology
Gothenburg, Sweden 2026

Finding representative routes of a vehicle's operations
Oliver Palada



Supervisor: Rickard Andersson, Volvo Group Trucks Technology
Supervisor: Carl Emvin, Department of Mechanics and Maritime Sciences
Examiner: Fredrik Bruzelius, Department of Mechanics and Maritime Sciences

Master's Thesis 2026
Department of Mechanics and Maritime Sciences
Division of Vehicle Engineering and Autonomous Systems
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone +46 31 772 1000

Cover: A truck driving a representative route. Generated using OpenAI ChatGPT.

Typeset in LaTeX
Gothenburg, Sweden 2026

Finding representative routes of a vehicle's operations
Oliver Palada
Department of Mechanics and Maritime Sciences
Division of Vehicle Engineering and Autonomous Systems
Chalmers University of Technology

# Abstract

Vehicles can be tested and assessed in many ways, such as simulation or in real traffic. Regardless of the assessment method, a test route must be defined, and, for the results to be meaningful, it should represent a realistic operating scenario. This thesis presents a method for identifying real-world test routes that are representative of a customer's road-related operating conditions.

Fixed driving cycles are commonly used in vehicle evaluation, but they may fail to capture the variation in road environments caused by customer-specific usage. This can reduce the accuracy of estimates of vehicle performance and energy consumption. To address this, customer log data are matched with road data and used to construct deterministic operating conditions. These are then used to estimate customer-specific stochastic operating condition parameters for road type, speed limits, slopes, curvature, and traffic control devices.

Routes are generated from a chosen origin and processed using the same road-data extraction procedure. A representativeness score is then defined as the degree to which the features of each route match the corresponding customer feature distributions. The score enables both generated routes and customer log files to be ranked according to how closely they match the customer's typical road operating conditions.

The results show that the proposed score can identify routes that are more representative of a customer's road usage, as well as routes corresponding to more unusual missions. Comparisons with estimated specific energy consumption indicate that the representativeness score is related to energy consumption, although it is not designed to optimise for it directly. A regression-based energy prediction model further suggests that the stochastic parameters describing the road contain relevant information for estimating energy consumption.

The thesis concludes that customer usage can be described by stochastic operating conditions and that this representation provides information that can support the statistical selection of representative real-world routes.

Keywords: operating conditions, stochastic processes, road modelling, route selection, vehicle energy consumption.

# Acknowledgements

I would like to express thankfulness to my examiner, Fredrik Bruzelius, for his feedback, questions, valuable ideas, and time throughout the project. His guidance helped improve both the direction and quality of the work.

I am also grateful to my supervisor, Carl Emvin, for his continuous support, many ideas, and time spent discussing the project. His input has been valuable in shaping and developing the thesis.

I would also like to thank my industrial supervisor, Rickard Andersson, at Volvo Group Trucks Technology, for welcoming me to Volvo and for providing valuable guidance, support, and encouragement throughout the thesis work.

Finally, I would like to thank everyone who has contributed with discussions, feedback, and support during the project.

Oliver Palada, Gothenburg, June 2026

# List of Acronyms

Below is the list of acronyms that have been used throughout this thesis:

| | |
|---|---|
| API | Application Programming Interface |
| AR(1) | First-order autoregressive process |
| CSV | Comma-separated values |
| dOC | Deterministic Operating Conditions |
| GTA | Global Transport Applications |
| OC | Operating Conditions |
| RMS | Root-mean-square |
| SEC | Specific Energy Consumption |
| sOC | Stochastic Operating Conditions |
| VECTO | Vehicle Energy Consumption Calculation Tool |

# Nomenclature

Below is the nomenclature for indices, sets, parameters, and variables used throughout this thesis.

## Indices

| | |
|---|---|
| $k$ | Sampled spatial index along a route |
| $m$ | Index for transition occurrence or traffic control device |
| $i, j$ | Indices for road types |
| $a, b$ | Indices for speed limits within a given road type |
| $f$ | Index for features |

## Sets

| | |
|---|---|
| $\mathcal{OC}_b$ | Bird's-eye view operating conditions representation |
| $\mathcal{OC}_s$ | Stochastic operating conditions representation |
| $\mathcal{OC}_d$ | Deterministic operating conditions representation |
| $\mathcal{S}_R$ | State space of road types |
| $\mathcal{S}_{V\|r_i}$ | State space of speed limits for road type $r_i$ |
| $\mathcal{F}$ | Set of features used in one score |

## Parameters

| | |
|---|---|
| $\lambda_{R_i}$ | Intensity parameter for road type transitions |
| $L_{R_i}$ | Mean segment length within road type $r_i$ |
| $p_{Rij}$ | Transition probability from road type $i$ to $j$ |
| $\boldsymbol{G}_R$ | Generator matrix for the road type process |
| $\boldsymbol{\pi}_R$ | Stationary distribution of road types |

| | |
|---|---|
| $\lambda_{V_a\|r_i}$ | Intensity parameter for transitions from speed limit $v_a$ within road type $r_i$ |
| $L_{V_a\|r_i}$ | Mean segment length for speed limit $v_a$ within road type $r_i$ |
| $p_{V_{ab}\|r_i}$ | Transition probability from speed limit $v_a$ to $v_b$ within road type $r_i$ |
| $\boldsymbol{G}_{V\|r_i}$ | Generator matrix for the speed limit process within road type $r_i$ |
| $\boldsymbol{\pi}_{V\|r_i}$ | Stationary distribution of speed limits within road type $r_i$ |
| $\phi_{Y\|r_i}$ | Autoregressive coefficient for road slope within road type $r_i$ |
| $e_{Y,k}$ | Innovation/noise term in the AR(1) slope model at index k |
| $\sigma^2_{e\|r_i}$ | Variance of noise in slope model within road type $r_i$ |
| $\mu_{Y\|r_i}$ | Mean road slope within road type $r_i$ |
| $\sigma^2_{Y\|r_i}$ | Variance of road slope within road type $r_i$ |
| $L_{h\|r_i}$ | Mean hill length within road type $r_i$ |
| $\Delta s$ | Sampling length |
| $\lambda_{C\|r_i}$ | Intensity of curve occurrences |
| $\mu_{C\|r_i}$ | Mean of log curvature distribution within road type $r_i$ |
| $\sigma^2_{C\|r_i}$ | Variance of log curvature distribution within road type $r_i$ |
| $\mu_{L\|r_i}$ | Mean of log curve length within road type $r_i$ |
| $\sigma^2_{L\|r_i}$ | Variance of log curve length within road type $r_i$ |
| $r_{\text{turn}}$ | Minimum turning radius constant |
| $\lambda_{E\|r_i}$ | Intensity of traffic control device of type $E$ |
| $p_{E\|r_i}$ | Probability of traffic control device occurrence of type $E$ on road type $r_i$ |
| $\varepsilon_{c\|r_i}$ | Curvature threshold within road type $r_i$ |
| $\mu_f$ | Mean value of feature f in the population |
| $\sigma_f$ | Standard deviation of feature f in the population |
| $\mu_{\text{ratio},i}$ | Mean distance ratio of road type $i$ |
| $\varepsilon_f$ | Offset for feature $f$ used in logarithmic transformation |
| $w_{\text{ratio}}$ | Weight applied to the road type ratio score |
| $w_i$ | Road type specific weights |

# Variables

| | |
|---|---|
| $R_k$ | Road type at index $k$ |
| $V_k$ | Speed limit at index $k$ |

| | |
|---|---|
| $X_m$ | Spatial position of the $m$:th traffic control device or transition |
| $Y_k$ | Road slope at the sampled index $k$ |
| $\tilde{Y}_k$ | Road slope at index $k$, including mean slope grade |
| $C_k$ | Road curvature at the sampled index $k$ |
| $C_m$ | Representative curvature of the $m$:th curve segment |
| $L_m$ | Length of $m$:th curve segment |
| $R'_m$ | Modified road radius of the $m$:th curve segment |
| $x_f$ | Value of feature $f$ for a candidate route |
| $\tilde{x}_f$ | Log-transformed value of feature $f$ |
| $z_f$ | Normalised deviation ($z$-score) of feature $f$ |
| $S_{\mathcal{F}}$ | RMS representativeness score over feature set $\mathcal{F}$ |
| $S_{\text{total}}$ | Total representativeness score |
| $S_i^{\text{penalty}}$ | Penalty applied when road type $i$ is missing |
| $S_{\text{energy}}$ | Energy score |

# Contents

# List of Figures

# List of Tables

# 1
# Introduction

This thesis applies the operating conditions (OC) format to the problem of finding representative routes. The chapter begins by describing current approaches for representing vehicle operations and their limitations, and then presents the problem statement, research questions, and ethical considerations.

## 1.1 Background

As the Earth becomes warmer, many legislative and executive bodies have set carbon dioxide emission goals to reduce the climate impact from transportation. To assess whether these goals are met, the EU Commission has developed a method for quantifying $CO_2$ emissions called VECTO. It simulates the vehicle motion corresponding to a set of driving cycles, a speed against time representation that is said to represent driving in most heavy vehicle operations [4]. Driving cycles are fixed representations, meaning that they neglect differences in vehicle use among specific customers. Relying on such representations when designing vehicles and optimising their energy consumption might therefore lead to sub-optimal solutions [11].

Heavy vehicles are often used in applications with different characteristics, for example, urban, regional, or long-haul operations. Each application type comes with different combinations of road types, speed levels, slopes, curvature, stops, and mission lengths. Since these factors affect vehicle energy consumption and the suitability of a given vehicle, a generic description of vehicle use may be insufficient when selecting or configuring a truck for a specific customer [2]. An alternative to fixed driving-cycle representations has therefore been developed at Chalmers in tight co-operation with RISE, VTI, Volvo, Scania and Volvo Cars: the OC format.

A customer-specific description of the operating conditions shall instead give a better basis for evaluating vehicles during development. This motivates the use of OC descriptions that capture how a vehicle is used. The OC describes the environment in which a customer operates using deterministic and stochastic processes that model road, weather, traffic, and mission characteristics at different levels of detail [13, 14]. These models have been adopted in a number of studies, including quantifying energy consumption, but their use cases are in this thesis extended to the selection of test routes suitable for a given customer.

## 1.2 Problem motivating the project

Final confirmation of results regarding, e.g., quantifying energy consumption across different operating environments, is achieved through real-world testing. In the ideal scenario, each vehicle model would be tested across all combinations of roads, weather conditions, traffic densities, and cargoes. This is unfortunately not feasible, so testing is instead conducted on shorter selected test routes. This raises the question of how such routes should be selected. Which, in turn, motivates the development of a method to evaluate candidate test routes against customer-specific sOC, with the goal of selecting routes representative of the customer's operating conditions [13].

## 1.3 Research questions

The primary research question is:

- How can a route be selected from the real-world road network such that it reflects the typical operating conditions of a specific customer?

This is addressed through three sub-questions:

1. How can the stochastic parameters of a collection of transport operations be estimated, i.e., how can a customer-specific sOC be constructed from log data and road attributes such as speed signs, slopes, curvature, etc?
2. How can candidate real-world routes be generated and modelled for statistical comparison against the customer-specific sOC?
3. How can route representativeness be quantified in order to evaluate candidate routes and identify which route best shares the characteristics of a given sOC?

## 1.4 Limitations

Only road-related factors of the OC are considered. Traffic, weather, and mission conditions are excluded due to time limitations and modelling complexity. While these factors influence vehicle operation in practice, disregarding them is a methodological choice, documented to avoid misinterpretation of the results.

## 1.5 Ethical considerations

This thesis is carried out in collaboration with Volvo Group. The research is primarily methodological and does not involve direct interaction with end users. However, several ethical aspects are relevant.

By improving route selection for testing and validation to better reflect real-world vehicle usage, the need for extensive physical testing may be reduced, as, for example, vehicle energy consumption evaluation is improved. This aligns with the Volvo

Group's long-term sustainability goals for reduced fuel consumption, lower greenhouse gas emissions, and more efficient vehicle development processes [18].

The method could also support vehicle development by making it possible to better match truck model selection or configurations with how a customer actually uses the vehicle. In particular, the results may allow manufacturers to better match vehicle specifications to real usage patterns, which reduces unnecessary energy consumption by avoiding inefficient vehicle configurations. In this way, the methodology may contribute to reduced fuel consumption, lower emissions, and improved sustainability in the transport sector.

The work relies on vehicle log data, obtained by Volvo from its customers, consisting of GPS trace files. These data files are used to derive road usage patterns through map matching and to estimate the statistical properties of the OC. No personal information is required for the methodology, and individual drivers are not analysed or profiled. However, the log data are still sensitive, and for this reason, they are used only for the intended purposes and are not shared with any external sources or presented directly, as only the statistical properties are reported.

# 2
# Operating conditions

The OC framework was originally introduced by [13], which formalised a representation of transport missions consisting of a bird's-eye view, stochastic operating conditions, and deterministic operating conditions (dOC). All three parts serve different purposes in the modelling of transport missions, but share the property of being divided into four categories: road, weather, traffic, and mission [14]. The bird's eye view is the highest and simplest level of representation, categorising different transport missions. The sOC models the stochastic variation of transport missions, while a realisation of such is described by the dOC.

This chapter provides an overview of the OC models, with a focus on the road component.

## 2.1 Bird's-eye view

The bird's-eye view consists of metrics that divide transport tasks within four categories: road, weather, traffic, and mission. This is summarised by the equation

$$\mathcal{OC}_b = \{\mathcal{R}_b, \mathcal{W}_b, \mathcal{T}_b, \mathcal{M}_b\}, \tag{2.1}$$

where the metrics within the sets $\mathcal{R}_b, \mathcal{W}_b, \mathcal{T}_b$, and $\mathcal{M}_b$ serve as statistical indicators in the categories road, weather, traffic and mission, respectively, with an example of road metrics given in the following sub-section.

### 2.1.1 Road metrics

Different approaches could be taken when categorising, for example, road topography. One such approach was developed by Volvo as part of their Global Transport Applications (GTA) system, which divides the road topography into four categories: flat, predominantly flat, hilly, and very hilly. The criteria for each category are specified in Table 2.1, with very hilly defined as when the other criteria are not satisfied.

The GTA categorisation is not definite; different road topography categorisations could be defined depending on use cases and purposes, but the GTA system is presented as an example.

**Table 2.1:** GTA road topography category specifications [2].

| Category | GTA label | Gradient | % of road |
|---|---|---|---|
| flat | FLAT | $< 3\%$ | $> 98\%$ |
| predominantly flat | P-FLAT | $< 6\%$ | $> 98\%$ |
| hilly | HILLY | $< 9\%$ | $> 98\%$ |
| very hilly | V-HILLY | - | - |

## 2.2 Stochastic operating conditions

The sOC mathematically models the variation within the four transport categories, describing their statistical properties and thereby enabling statistical reproduction of transport operations. The full model is given by

$$\mathcal{OC}_s = \{\mathcal{R}_s, \mathcal{W}_s, \mathcal{T}_s, \mathcal{M}_s\}, \tag{2.2}$$

where $\mathcal{R}_s, \mathcal{W}_s, \mathcal{T}_s, \mathcal{M}_s$ are the sets containing the statistical properties for each parameter within the categories road, weather, traffic and mission, respectively.
Initially, the sOC consisted only of the road category, but was later extended by [14], which introduced parsimonious stochastic models for weather and traffic. The sOC representation was further extended in [15] and [3] regarding mission properties.

The road component of the sOC is divided into two hierarchical submodels to heuristically introduce parameter dependencies while preserving modularity and parsimony. The primary model describes road types, while the secondary model describes road properties such as curvature, slope, and speed limit, with parameters estimated separately for each road type.

### 2.2.1 Road type model

It is not well defined what a road type is, as it cannot be directly observed or measured. In this work, as in previous work [13, 14], speed limits are used as an indicator of road type. Using this method, a route is assumed to consist of segments of a finite number of distinct categories. These categories are ordered and separated by speed limit thresholds, so that moving from one threshold to the next corresponds to a change in road type. Each speed limit value is therefore mapped to exactly one road type, as specified in Table 2.2.

Under this formulation, the road type is treated as a discrete random variable $R_k$ taking values in the state space $\mathcal{S}_R = \{r_1, ..., r_{n_r}\}$, depending on the sequence of speed limits encountered along a route, where $n_r$ is the number of road type categories. In this work, the set $\mathcal{S}_R = \{r_1, r_2, r_3\}$, where $r_1 =$ Urban, $r_2 =$ Rural, $r_3 =$ Highway, is used. The process is assumed to follow a first-order Markov process [1], meaning that the probability of the next road type depends only on the current road type:

$$\mathbb{P}(R_{k+1} = r_j \mid R_k = r_i, R_{k-1}, \dots, R_1) = \mathbb{P}(R_{k+1} = r_j \mid R_k = r_i). \tag{2.3}$$

**Table 2.2:** Speed limit to road type mapping.

| Speed limit | Road type |
|---|---|
| 30<br>40<br>50 | Urban |
| 60<br>70<br>80 | Rural |
| 90 | Highway |

In addition, the spatial positions $X_m$ at which road type changes occur are modelled as a Poisson process [1]. Therefore, the distances between successive transitions are exponentially distributed:

$$X_{m+1} - X_m \sim \mathrm{Exp}(\lambda_{R_i}), \tag{2.4}$$

where the intensities are calculated from $L_{R_i}$, the mean segment length for road type $R_i$, as

$$\lambda_{R_i} = \frac{1}{L_{R_i}}. \tag{2.5}$$

The transitions between road types are modelled using a Markov chain, with transition probabilities estimated from the observed transition counts in the road data. The stationary distribution $\boldsymbol{\pi}_R$ of the road type process represents the long-term fraction of distance travelled on each road category, and is calculated as

$$\boldsymbol{\pi}_R \boldsymbol{G}_R = 0, \tag{2.6a}$$

$$\sum_{i=1}^{n_r} \pi_{R_i} = 1. \tag{2.6b}$$

The generator matrix $\boldsymbol{G}_R$ with entries $g_{Rij}$ is given by

$$g_{Rij} = \begin{cases} \lambda_{Ri} p_{Rij}, & i \neq j, \\ -\lambda_{Ri}, & i = j, \end{cases} \tag{2.7}$$

where $p_{Rij}$ are the transition probabilities from road type $r_i$ to $r_j$ and self-transitions are not allowed, i.e. $p_{Rii} = 0$, which implies that

$$\sum_{j=1}^{n_r} p_{Rij} = 1, \quad i = 1, \dots, n_r,\ i \neq j. \tag{2.8}$$

The parameters necessary for the road type model are the mean segment lengths $L_{R_i}$ and the transition matrix $\boldsymbol{P}_R$, with entries $p_{Rij}$, since all other parameters can be calculated from them.

### 2.2.2 Speed limit model

The speed limit model describes the sequence of speed limits along a route. Since the speed limit model is a secondary model, its parameters are calculated separately for each road type. For a given road type $r_i$, the speed limit is treated as a discrete random variable $V_k$ taking values in a finite state space $\mathcal{S}_{V|r_i} = \{v_1, \ldots, v_{n_v|r_i}\}$, where $v_1, \ldots, v_{n_v|r_i}$ are the speed limits within road type $r_i$, and $n_v \mid r_i$ is the number of speed limits in road type $r_i$. The speed limits for each road type, in this work, are given in Table 2.2.

The sequence of the speed limits along the route is modelled as a first-order Markov chain [1], such that

$$\mathbb{P}(V_{k+1} = v_b \mid V_k = v_a, R_k = r_i) = p_{V_{ab}|r_i}, \tag{2.9}$$

where $p^{(r_i)}_{V_{ab}}$ denotes the transition probability from speed limit $v_a$ to $v_b$ within road type $r_i$.

Similar to the road type model, the spatial positions $X_m$ at which speed limit changes occur are modelled as a Poisson process. Therefore, the distances between successive transitions are exponentially distributed:

$$X_{m+1} - X_m \sim \mathrm{Exp}(\lambda_{V_a|r_i}), \tag{2.10}$$

where the intensity is defined from the mean segment length $L_{V_a|r_i}$ estimated from the road data for speed limit $v_a$ and road type $r_i$ as

$$\lambda_{V_a|r_i} = \frac{1}{L_{V_a|r_i}}. \tag{2.11}$$

The stationary distribution $\boldsymbol{\pi}_{V|r_i}$ of the speed limit process represents the long-term fraction of distance travelled on each speed limit state within road type $r_i$, and is calculated as

$$\boldsymbol{\pi}_{V|r_i} \boldsymbol{G}_{V|r_i} = 0, \tag{2.12a}$$

$$\sum_{a=1}^{n_v|r_i} \pi_{V_a|r_i} = 1. \tag{2.12b}$$

The generator matrix $\boldsymbol{G}_{V|r_i}$ with entries $g_{V_{ab}|r_i}$ is given by

$$g_{V_{ab}|r_i} = \begin{cases} \lambda_{V_a|r_i} p_{V_{ab}|r_i}, & a \neq b, \\ -\lambda_{V_a|r_i}, & a = b, \end{cases} \tag{2.13}$$

where $p_{V_{ab}|r_i}$ are the transition probabilities from speed limit $v_a$ to $v_b$, conditioned on road type $r_i$, and self-transitions are not allowed, which implies that

$$\sum_{b=1}^{n_v|r_i} p_{V_{ab}|r_i} = 1, \quad a = 1, \ldots, n_{v|r_i}, \; a \neq b. \tag{2.14}$$

The speed limit model for each road type $r_i$ is fully described by the mean segment lengths $L_{V_a|r_i}$, and the transition matrix $\boldsymbol{P}_{V|r_i}$, where only transitions between allowed speed limits within each road type are considered.

### 2.2.3 Road topography model

Road topography describes the longitudinal gradient of the road. The gradient sequence is modelled as a first-order autoregressive process (AR(1)) [10], which is a stochastic process with spatial correlation.

Let $Y_k$ denote the gradient at spatial index $k$, conditioned on road type $r_i$:

$$Y_k = \phi_{Y|r_i} Y_{k-1} + e_{Y,k}, \quad e_{Y,k} \sim \mathcal{N}(0, \sigma^2_{e|r_i}). \tag{2.15}$$

The parameter $\phi_{Y|r_i} \in (-1, 1)$ is the autoregressive coefficient describing the spatial correlation of the slope. The innovation term $e_{Y,k}$, or white-noise at index k, represents the part of the gradient that is not explained by the previous slope value, and is assumed to be independent and normally distributed with zero mean and variance $\sigma^2_{e|r_i}$.

The autoregressive coefficient can also be related to a mean hill length $L_{h|r_i}$ for road type $r_i$. The relation is written as

$$\phi_{Y|r_i} = \cos\left(\frac{2\Delta s}{L_{h|r_i}}\right), \tag{2.16}$$

where $\Delta s$ is the sampling length. Equivalently,

$$L_{h|r_i} = \frac{2\Delta s}{\arccos(\phi_{Y|r_i})}. \tag{2.17}$$

The stationary variance $\sigma^2_{Y|r_i}$ of the slope process is given by

$$\sigma^2_{Y|r_i} = \frac{\sigma^2_{e|r_i}}{1 - \phi^2_{Y|r_i}}. \tag{2.18}$$

In practice, routes of limited length have a mean slope grade value $\mu_{Y|r_i}$, so that the slopes become

$$\tilde{Y}_k = Y_k + \mu_{Y|r_i}. \tag{2.19}$$

The parameters $\phi_{Y|r_i}$, $\sigma^2_{e|r_i}$, and $\mu_{Y|r_i}$ are estimated from the road data for each road type.

### 2.2.4 Road curvature model

Curves are treated as isolated stochastic events along the road. Each curve is described by $(X_m, C_m, L_m)$, representing the location, modified curvature, and length of the $m$:th curve segment.

The locations $X_m$ of curves are modelled as a Poisson process [1], and therefore the distances between curves are exponentially distributed:

$$X_{m+1} - X_m \sim \text{Exp}(\lambda_{C|r_i}), \tag{2.20}$$

where $\lambda_{C|r_i}$ is the intensity parameter describing the expected number of curves per unit of distance for road type $r_i$.

The curvature is modelled using a modified log-normal distribution:

$$\frac{1}{C_m} = R'_m + r_{\text{turn}}, \quad \ln R'_m \sim \mathcal{N}(\mu_{C|r_i}, \sigma^2_{C|r_i}), \tag{2.21}$$

where $R'_m$ is the modified road radius of the $m$th curve segment, $\mu_{C|r_i}$ is the mean curvature of road type $r_i$, $\sigma^2_{C|r_i}$ is the curve variance of road type $r_i$, and $r_{\text{turn}}$ is a constant representing the minimum turning radius of the road.

The curve length is also modelled as a log-normal distribution:

$$\ln L_m \sim \mathcal{N}(\mu_{L|r_i}, \sigma^2_{L|r_i}), \tag{2.22}$$

where $\mu_{L|r_i}$ is the mean length of curves, and $\sigma^2_{L|r_i}$ is the variance of curve lengths, both within road type $r_i$.

The parameters $\lambda_{C|r_i}$, $\mu_{C|r_i}$, $\sigma^2_{C|r_i}$, $\mu_{L|r_i}$, and $\sigma^2_{L|r_i}$ are estimated from the road data for each road type. In parameter estimation, curves are identified using the threshold $\varepsilon_{c|r_i}$, as explained in Section 3.1.5.

### 2.2.5 Road traffic control devices

Road traffic control devices such as traffic lights, stop signs, yield signs, and pedestrian crossings are modelled as stochastic point events along the road. For each control device type, the occurrence points are modelled as a Poisson process [1]. The distances between consecutive control devices are therefore exponentially distributed:

$$X_{m+1} - X_m \sim \text{Exp}(\lambda_{E|r_i}), \tag{2.23}$$

where $\lambda_{E|r_i}$ is the intensity parameter for traffic control device of type $E$ on road type $r_i$.

Equivalently, in a discretised spatial representation with step length $\Delta s$, the occurrence of a traffic control device at each step can be approximated as a Bernoulli trial with probability

$$p_{E|r_i} = \lambda_{E|r_i} \Delta s. \tag{2.24}$$

The intensity parameters $\lambda_{E|r_i}$ are estimated as the number of observed traffic control devices per unit distance in each road type.

## 2.3 Deterministic operating conditions

Similar to the two representations already discussed, the dOC is summarised by the equation

$$\mathcal{OC}_d = \{\mathcal{R}_d, \mathcal{W}_d, \mathcal{T}_d, \mathcal{M}_d\}, \tag{2.25}$$

where $\mathcal{R}_d, \mathcal{W}_d, \mathcal{T}_d, \mathcal{M}_d$ are the sets containing the dOC parameters for road, weather, traffic, and mission categories, respectively.

As the name implies, a specific dOC represents a vehicle's mission, providing details of an individual transport operation. This representation may be derived directly from log and road data, and the derived dOC may subsequently be used to estimate sOC parameters. In the opposite direction, dOC may also be generated as a realisation of sOC by sampling from the stochastic parameters. Then, after sampling dOC, they can be used in simulations to examine how, e.g. energy consumption is affected by stochastic parameter values within the different categories.

# 3
# Method

Stochastic models and parameter estimation are used to describe both a customer's operating conditions and generated routes from road data, with the goal of selecting representative routes in a real-world road network. This chapter describes the methodology used, from route generation and parameter estimation to route evaluation and selection. The chapter ends with a discussion of the proposed method's modelling choices and limitations.

## 3.1 Route selection methodology

The route selection process starts from either customer vehicle log data or generated routes. Data are processed through route matching and road-data extraction to construct dOCs, which are subsequently used to estimate sOC parameters. Candidate routes are evaluated against the customer-specific sOC using a representativeness score to select an optimal test route. All program code is written in the programming language Python.

Candidate routes are generated using a combination of the HERE Isoline Routing API [6] and the HERE Routing API [8]. Trace files are processed using the HERE Route Matching API [7], with relevant road attributes used to build the dOC extracted from the resulting route. The dOC data is interpolated to obtain equidistant samples along the route, simplifying the estimation of the sOC parameters. Lastly, routes are evaluated against customer-specific sOC using a representativeness score. The method's final output is a numerical ranking of how well each route reflects a customer's operating conditions. The workflow is illustrated in Figure 3.1.

### 3.1.1 Route generation

The purpose of route generation is to create a set of routes with similar total distance, starting from a predefined origin, so they can be used as candidates for test route selection.

An isoline is generated around the origin using truck mode, fast routing, and a specified distance. An isoline refers to a boundary of points on a map that can be reached from the origin within a given distance. The returned isoline is encoded as a flexible polyline [5], which is decoded into latitude and longitude coordinates. A chosen number of points are then sampled along the isoline and used as endpoints.

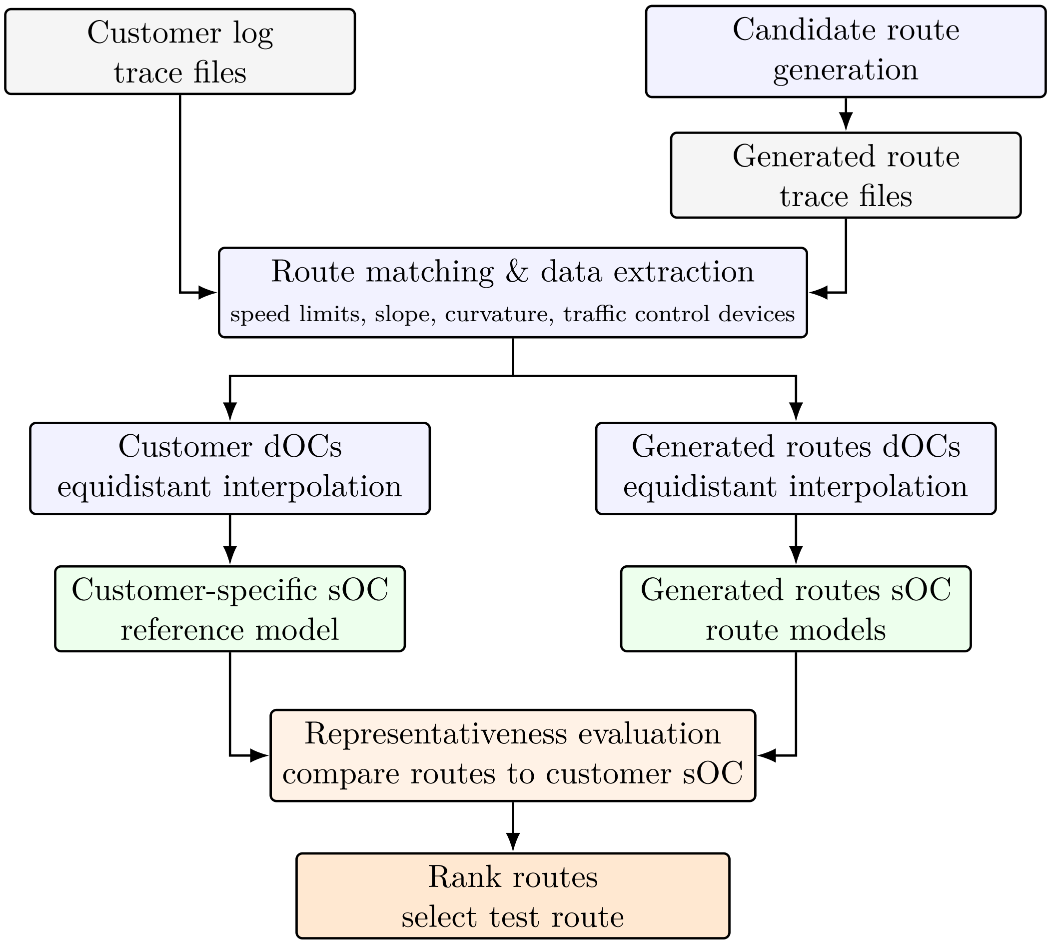



**Figure 3.1:** Overview of the method for selecting representative test routes. Customer log data and generated routes are processed through route matching with road data extraction to construct dOCs, estimate sOC parameters, evaluate representativeness, and select a route.

A separate route is requested to each endpoint from the origin, using the HERE Routing API with truck and fast mode. The returned route is again decoded from the flexible polyline and stored as a trace file containing latitude and longitude coordinates. The route length is then estimated from the coordinates, and routes with estimated length outside of a specified tolerance of the target distance are discarded. The generated candidate set then consists of routes that start at the same origin and have approximately the desired total distance. The sampled isoline endpoints are also saved separately for reproducibility.

A second route generation script was developed to manually create routes from a user-specified origin and destination, optionally including via points. It requests both the fastest and shortest routing alternatives and stores both if they differ. This script is not used for the main route generation process, but is useful when a specific route, destination, or road segment needs to be analysed.

### 3.1.2 Map generation

Routes are visualised on interactive maps using a script that generates HTML maps from OpenStreetMap [12], showing the route, approximate distance, and its start and end points with markers. The purpose is to inspect, for example, the best-scoring routes and determine whether they are suitable for use as test routes, or to view the routes of the customer log files.

### 3.1.3 Construction of a dOC

A customer dOC CSV file is created from each trace file in the collection of customer log files, similarly, a candidate route dOC CSV file is created for each generated trace file using the same script. All log files must include the latitude and longitude coordinates, which are map-matched to the roads using the HERE Route Matching API. Route matching enables the extraction of road attributes along a route by mapping road segments and retrieving their corresponding attributes. The API response is preprocessed in order to obtain a consistent sequence of road attributes. The script extracts distance, speed limit, slopes, curvatures, and traffic control device indicators for traffic lights, stop signs, yield signs, and pedestrian crossings. Each dOC file represents a route and describes the road properties encountered along it.

### 3.1.4 Equidistant discretisation

The sOC are defined in continuous space, but for implementation purposes, the dOC files are discretised into equidistant spatial steps. This is necessary because the HERE Route Matching API response contains points with varying distances between them. In this work, a fixed sample length of $\Delta s = 25$ m is used. Let $\Delta s$ denote the sample length, the equidistant spatial sample points are defined as

$$s_k = k\Delta s, \tag{3.1}$$

where $k$ is the sample index and $s_k$ the distance from the start of the route. Boolean traffic control device indicators are assigned to the closest point on the equidistant grid, while other road attribute values are interpolated. Speed limit values are interpolated using nearest-neighbour interpolation, and numeric variables are linearly interpolated. Equidistant discretisation simplifies numerical implementation while preserving the statistical properties of the continuous models.

### 3.1.5 Estimation of sOC parameters

From the equidistant dOC CSV files, statistical parameters describing the road conditions are estimated. These parameters include distributions of road types, transition rates between road types, slope and curvature statistics, and traffic control device intensities, in accordance with the sOC framework. The secondary parameters are estimated separately for each road category, urban, rural and highway. Speed limits of 10 and 20 km/h are treated as outliers and therefore recoded to 30 km/h before building the sOC, since they occur only over negligible distances and are not expected to considerably affect the results. The resulting parameter set defines the sOC, which describes the statistical road properties of the customer's missions.

Curve segments are identified from the curvature values using a threshold $\varepsilon_{c|r_i}$, defined for each road type $r_i$. A sample at position $k$ is classified as a curve if

$$|C_k| > \varepsilon_{r_i}, \tag{3.2}$$

where $C_k$ is the curvature value. The threshold is calculated using a representative speed level $v_{r_i}$ for road type $r_i$ as

$$\varepsilon_{c|r_i} = \frac{a_{\text{ref}}}{(v_{r_i}/3.6)^2}, \tag{3.3}$$

where $v_{r_i}$ is given in km/h and chosen as $v_{r_i} = 50$, $v_{r_i} = 80$ and $v_{r_i} = 90$ for urban, rural and highway. The reference lateral-acceleration level, $a_{\text{ref}}$, is, for simplicity, chosen as $a_{\text{ref}} = 0.5$. The formula implies that the curvature threshold decreases with increasing speed.

Consecutive samples satisfying $|C_k| > \varepsilon_{c|r_i}$ are grouped into curve segments, from which the parameters of the curvature model are estimated. Segments shorter than 50 m are discarded before estimating curvature statistics. For each curve segment, the curvature is defined as the mean absolute curvature over the samples belonging to that segment. In the parameter estimation, the minimum turning radius constant is fixed at $r_{\text{turn}} = 10\,\text{m}$ for all road types.

### 3.1.6 Representativeness score

Route representativeness is quantified by comparing the road characteristics of candidate routes to those of the customer-specific sOC. The customer's sOC defines a probabilistic reference model, while candidate routes are treated as deterministic with attributes evaluated against this model. Evaluation is done using an objective function. The representative route selection problem can then be formulated as an optimisation problem, aiming to identify a route that minimises the representativeness objective function. Different approaches can be used to define the objective function, with those used in this thesis described in detail in Chapter 4.

### 3.1.7 Energy consumption estimation

Energy consumption was estimated for each log file using an unpublished energy estimation script.[1] Only the assumptions relevant to this thesis are described.

Weather-related input variables were set to constant values for all log files. Ambient temperature was set to 10 °C, and wind speed to 0.1 m/s. In addition, no mission stops were included. The resulting estimates should therefore be interpreted as specific energy consumption (SEC) values under the fixed weather and without mission stops.

[1] The script was provided by Carl Emvin and was developed as part of ongoing research.

## 3.2 Method discussion and limitations

The method has several limitations and modelling choices that should be considered when interpreting the results.

The generated routes tend to use faster roads because the routing uses fast mode. Depending on the customer data, starting point, and isoline distance, many generated routes may not be representative of the road type distribution. Also, sampling endpoints along the isoline means that the generated route set depends on the number of sampled points and does not cover all possible routes of the desired length.

The customer-specific sOC is estimated after combining vehicle trajectory data with road data through HERE Route Matching. The quality of the results depends on the underlying road data and the extraction of road attributes from the route matching response. A modelling choice in this work is to restrict the representation of the highway speed limit to the speed levels observed via HERE Route Matching in truck mode, meaning that the highest highway speed limit in the analysed data is 90 km/h. In the original operating conditions formulation, the speed limit model is based on legal speed signs, resulting in a larger set of speed limits. The decision was made to avoid errors when using truck mode for both the HERE Routing API and the HERE Route Matching API, but this reduces comparability with implementations based on speed-sign representation.

The secondary models for slope, curvature and traffic control device intensities are estimated by filtering the full equidistant route data by road type and then pooling the resulting samples within each category. This means that samples from the same road type are combined even when they come from different parts of the route. If a road type transition occurs while the vehicle is travelling on a slope or in a curve, the feature is split at the transition, and continuity is not preserved. The approach is computationally convenient, but it may affect the estimated topography and curvature statistics when features span over multiple road types.

# 4

# Quantification of route representativeness

## 4.1 Representativeness score

A representativeness score is introduced to evaluate how well a route reflects the customer-specific sOC. The representativeness score is based on comparing statistical features of a route to the corresponding distributions derived from customer data. A feature can be a parameter, a log-transformed parameter, or a derived route statistic such as road type ratio or average speed.

The features are standardised before aggregation, since the selected road features have different units and levels of variability. Scaling variables before combining them is a method commonly used in multivariate statistical analysis [9], as some variables may otherwise dominate the resulting scores. For each feature, a normalised deviation is computed using a $z$-score. Let $x_f$ denote the value of feature $f$ for a route, and let $\mu_f$ and $\sigma_f$ denote the mean and standard deviation of that feature in the population (collection of customer data). The normalised deviation, or $z$-score, is then given by

$$z_f = \frac{x_f - \mu_f}{\sigma_f}, \tag{4.1}$$

for feature $f$ from the selected feature set. A $z$-score close to zero means that the route feature is close to the population mean, relative to the variation in the population.

To aggregate feature deviations into a single non-negative score, a root-mean-square (RMS) measure is used. For a feature set $\mathcal{F}$, the score is defined as

$$S_{\mathcal{F}} = \sqrt{\frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} z_f^2}. \tag{4.2}$$

The RMS score is computed separately for four feature groups: road type ratios, urban features, rural features, and highway features. Denoting these feature sets as ratio, urban, rural and highway, the total score is defined as

$$S_{\text{total}} = w_{\text{ratio}} S_{\text{ratio}} + w_{\text{urban}} S_{\text{urban}} + w_{\text{rural}} S_{\text{rural}} + w_{\text{highway}} S_{\text{highway}}, \tag{4.3}$$

where $S_{\text{ratio}}$ measures the deviation in road type ratio distribution, and $S_{\text{urban}}$, $S_{\text{rural}}$, and $S_{\text{highway}}$ are the scores of the per road type feature sets. Equivalently, this can be written as

$$S_{\text{total}} = w_{\text{ratio}} S(\mathcal{F}_{\text{ratio}}) + \sum_{i=1}^{n_r} w_i S_{\mathcal{F}}, \qquad \mathcal{F} = \text{urban, rural, highway}, \tag{4.4}$$

where $\mathcal{F}$ denotes the feature set of each road type, $w_i$ the corresponding road type-specific weight, and $n_r$ is the number of road types. The road type-specific weights are defined using mean road type ratios estimated from the population data,

$$w_i = \mu_{\text{ratio},i}, \qquad i = 1, \ldots, n_r, \tag{4.5}$$

where $\mu_{\text{ratio},i}$ is the mean distance ratio of road type $i$. These weights ensure that scores of road types that occur more frequently in the customer data contribute more to the total score.

### 4.1.1 Feature set definition

The feature sets $\mathcal{F}$ used in the RMS scores are mainly constructed from parameters available in the sOC. However, only a subset of the available parameters is included in the representativeness score. The features are divided into four groups: road type ratios and road type-specific features for urban, rural, and highway roads.

The road type ratio feature set consists of

- Road type ratios: urban ratio, rural ratio, highway ratio,

which describe the fraction of distance travelled on each road type.

For urban and rural road types, the following features are included:

- Traffic control device intensities: traffic lights, stop signs, yield signs, and pedestrian crossings,
- Speed characteristics: average speed,
- Topography: mean slope grade, variance of slope grade, and mean hill length,
- Curvature: curve intensity, mean curvature radius, variance of curvature radius, and mean curve length.

Some features are log-transformed before evaluation to make $z$-score normalisation more meaningful for strongly skewed variables. The exact transformed features and the offset choice are described in Section 4.1.2.

For highway roads, a reduced set of features is used:

- Topography: mean slope grade, variance of slope grade, and mean hill length,
- Curvature: curve intensity, mean curvature radius, variance of curvature radius, and mean curve length.

Traffic control device intensities and average speeds are excluded for highways because they are often constant, leading to little or no variance and therefore unstable $z$-scores.

Some estimated sOC parameters are not included in the feature sets. In particular:

- Transition probabilities and generator matrix parameters are excluded, as the representativeness score is based on aggregated statistical properties rather than sequences,

- Parameters with negligible variance are excluded to avoid numerical instability in the $z$-score computation,
- Some features are judged to be redundant or highly correlated and therefore avoided by using a limited set of summary statistics.

The resulting feature sets should provide a balance between capturing road characteristics and maintaining a numerically stable result.

### 4.1.2 Feature log-transformation

Some of the estimated parameters for road features are strictly non-negative and right-skewed, in particular, the traffic control device intensities. Using such values directly in Eq. 4.1 leads to $z$-scores that are affected by the skewness rather than by representative deviation from the customer population.

Although z-score standardisation does not require normally distributed features, its interpretation as a deviation from a typical value is more meaningful when the underlying feature distributions are not strongly skewed. To reduce this effect, a logarithmic transformation is applied to the problematic features. The transformed value of a feature $x_f$ is defined as

$$\tilde{x}_f = \log(x_f + \varepsilon_f), \tag{4.6}$$

where $\varepsilon_f > 0$ is a feature-specific offset. The transformed feature values are then used when computing the population mean $\mu_f$, standard deviation $\sigma_f$, and route-specific feature value $x_f$ in the $z$-scores.

The following features are log-transformed in the representativeness score:

- Urban: traffic light intensity, stop sign intensity, pedestrian crossing intensity, variance of slope grade, mean hill length.
- Rural: traffic light intensity, stop sign intensity, yield sign intensity, pedestrian crossing intensity, slope grade variance.
- Highway: slope grade variance, mean hill length.

The choice of which features to transform was based on inspection of their distributions. Features were log-transformed when the transformation reduced skewness and made the distributions closer to normal or symmetric distributions, thereby improving their suitability for $z$-score calculations.

Since the features to be transformed can attain zero values, the positive offset $\varepsilon_f$ in Eq. (4.6) is required, which is determined separately for each transformed feature using the customer data. Let $x_f^+$ denote the strictly positive observations of feature $f$ in the population. The offset is then defined as

$$\varepsilon_f = \frac{1}{2} Q_{0.05}(x_f^+), \tag{4.7}$$

where $Q_{0.05}(x_f^+)$ denotes the 5% quantile of the strictly positive observations. If fewer than five positive observations are available, the offset is instead set to one-half of the smallest positive observed value. This choice provides a small offset relative to the scale of each feature, while avoiding undefined logarithms at zero. The same $\varepsilon_f$ values are then used for both the population parameters and the generated routes' feature values, ensuring that the offset values are not re-adapted to the generated routes, as this would affect the $z$-scores. Log-transformations are used to stabilise variance since several transformed features contain many zero observations. The offset definition in Eq. 4.7 was chosen to make the logarithmic transformation well-defined while keeping the offset small relative to the scale of the observations. No extensive analysis was performed to select the offset, and other offset definitions or transformations could be considered.

The distributions before transformation are shown in Section 5.2, and after transformation in Appendix A, indicating that the selected transformations make the distributions more suitable for use in $z$-scores.

### 4.1.3 Penalty for missing road types

The representativeness score includes a penalty for missing road types in evaluated routes. Without the penalty, a route could obtain a low score when missing a road type by avoiding deviations in its associated features, resulting in $S_i = 0$, where $S_i$ is the corresponding RMS score for road type $i$.

To prevent this, a penalty is introduced whenever a road type $i$ is absent in the candidate route (i.e., its distance ratio is zero) but is present in the population. The penalty is defined as the $z$-score corresponding to a zero road type ratio value:

$$S_i^{\text{penalty}} = \left| \frac{0 - \mu_i}{\sigma_i} \right|, \tag{4.8}$$

where $\mu_i$ and $\sigma_i$ are the mean and standard deviation of the road type ratio in the population. $S_i$ is then replaced with $S_i^{\text{penalty}}$ in Eq. 4.3.

This ensures that missing a road type is penalised proportionally to how common that road type is in the customer data. In particular, if a road type has a large expected ratio, its absence will result in a larger penalty. However, the penalty serves as a safeguard in the ranking, and the exact numerical value is not of importance. Its main purpose is to prevent routes that completely omit a road type present in the customer data from being selected.

### 4.1.4 Interpretation and discussion

The full representativeness scoring process is illustrated in Figure 4.1. A customer's feature distributions are used to estimate the reference statistics. The corresponding feature values are then extracted for each candidate route and compared to the customer reference distributions using z-scores. These deviations are aggregated separately for the road type ratios and for each road type-specific feature group using RMS scores. The group scores are then combined using the chosen weights

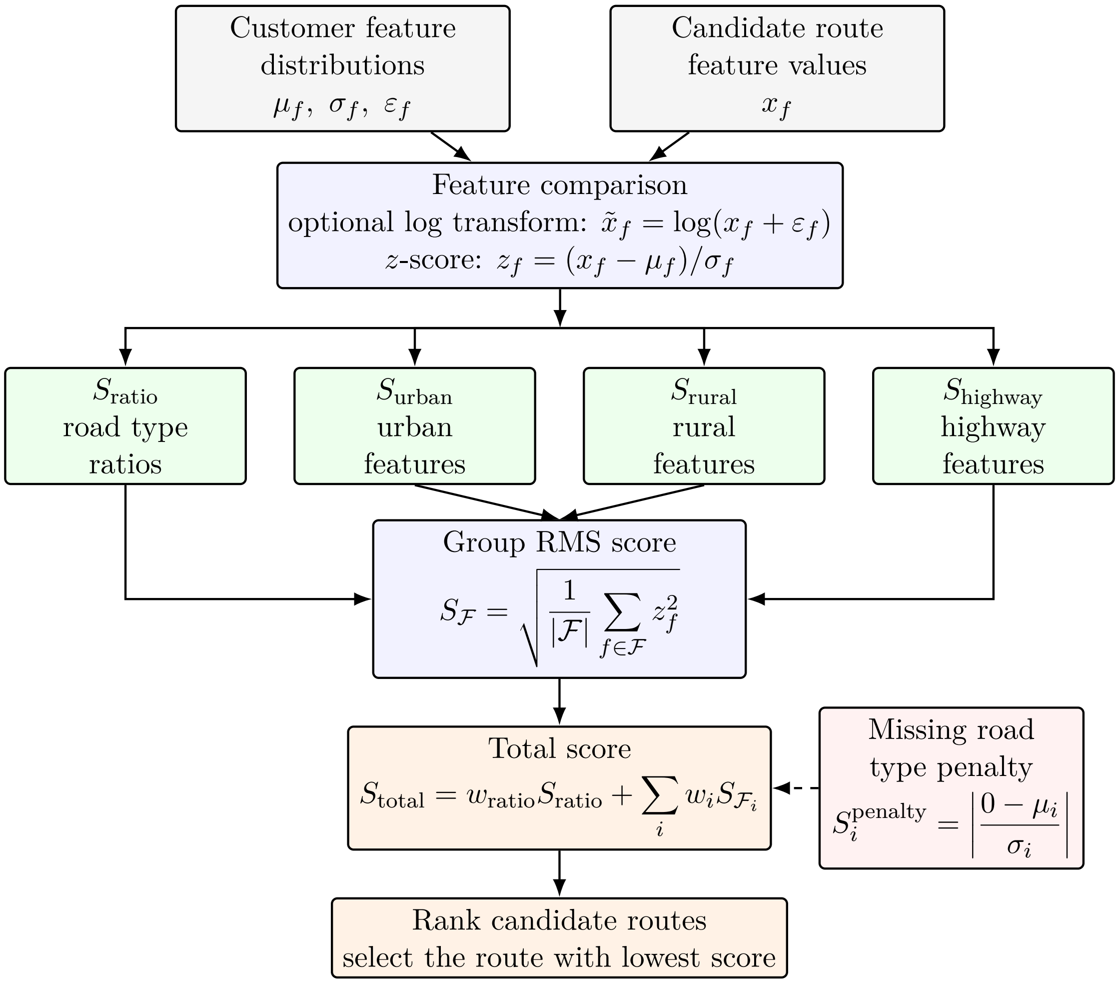



**Figure 4.1:** Overview of the representativeness scoring. Customer feature distribution parameters are estimated, candidate route features are standardised, aggregated within feature groups, and then combined into a total score. Optionally, the score is adjusted for missing road types.

to obtain the total representativeness score. If a candidate route lacks a road type present in the customer data, the corresponding road type score is replaced with a penalty before the final ranking is performed.

The $z$-score formulation is used because it provides a measure of how far a specific route deviates from the customer population with respect to the selected road features. Parameters with different units and scales can be combined into a single score, since each feature is normalised by its population standard deviation. The RMS formulation is used because it produces a single non-negative measure, since squaring the z-scores prevents positive and negative deviations from cancelling each other. RMS also gives larger penalties to features that deviate strongly from the customer population. This is desirable, since a route that is highly unrepresentative in one road characteristic should receive a higher score even if it is close to the customer mean for several other characteristics. The resulting total score is also interpretable: a lower score indicates that the candidate route is closer to the customer mean relative to the variation in the customer data.

The separate RMS scores for road type ratios and for each road type make the scoring modular. Instead of aggregating all features into a single RMS score directly, the formulation keeps the contributions separate until the final weighted sum. This makes the score easier to interpret and allows weighting to be modified depending on the intended application. For example, weights can be adjusted if certain road environments are more relevant for a particular customer or vehicle evaluation.

The influence of road type is significant, as it serves as the primary model in the sOC framework, and the secondary parameters vary by road type. This is reflected in the ratio score $S_{\text{ratio}}$ in Eq. 4.3, with weight $w_{ratio} = 1.0$, which corresponds to the total weight of the other components. As a result, deviations in road type distribution have a relatively large impact on the total score, which reflects the importance of road type, but this effect could still be over- or underestimated. The purpose of weighting road type-specific components is to reflect that deviations in the parameters of a dominant road type should have a larger impact on the score.

All road features of the secondary models contribute equally to the score of the corresponding road type. In practice, different parameters are expected to have different impacts on, e.g. vehicle energy consumption, which this formulation does not capture. Also, some characteristics are represented by multiple parameters. For example, curvature is described by five parameters, of which four are included in the RMS score, which may lead to overweighting of curvature compared to other features described by only one parameter.

The highway road type has fewer parameters included in the RMS score than urban and rural road types. The reason for this is that some features are constant on highways, such as intensities for e.g. stop signs. This implies that each included parameter has a larger influence on the score for highways, as $|\mathcal{F}|$ is smaller. While this introduces an imbalance, it is in part desirable, since the scores $S_{\mathcal{F}}$ for different road types should remain comparable on the same scale.

Finally, it should be noted that the representativeness score is not designed to optimally predict energy consumption. Energy consumption is used later for comparison and analysis to evaluate whether routes that are statistically representative of sOC road features are also representative of estimated energy consumption.

## 4.2 Energy score

If the energy estimator as described in Section 3.1.7 is available, the estimated SEC of routes can be compared with the customer population. This is done by defining an energy score as the absolute $z$-score of the estimated SEC,

$$S_{\text{energy}} = \frac{|x_{\text{SEC}} - \mu_{\text{SEC}}|}{\sigma_{\text{SEC}}}, \tag{4.9}$$

where $\mu_{\text{SEC}}$ and $\sigma_{\text{SEC}}$ are the mean and standard deviation of the estimated SEC values in the customer data. A low energy score therefore means that the route has an estimated energy consumption per travelled distance close to the customer average.

The energy score can be used directly to rank routes by estimated energy consumption, or possibly in combination with the representativeness score, as

$$S_{\text{total,energy}} = S_{\text{total}} + w_{\text{energy}} S_{\text{energy}}. \tag{4.10}$$

However, this is not always desirable if the goal is to select a route representative of the road conditions. Different road attributes may compensate for each other in the energy estimate. For example, a route may have higher energy demand from topography but lower energy demand from speed or curvature, resulting in a representative energy score even though the road characteristics are not representative. For this reason, the energy score is not used directly in this thesis, but rather serves as a basis for comparison with the road-feature-based representativeness score.

## 4.3 Energy prediction score

The energy score in Section 4.2 requires that the SEC is estimated for each evaluated route. Since this requires the energy-estimation script, a second score is introduced to predict the energy score using the road features used in the representativeness score. The purpose is to investigate whether a regression model can learn an energy-related weighting of road features.

For each route $n$, the same signed $z$-scores as in Eq. 4.1 are computed for the selected regression feature set $\mathcal{F}_{\text{reg}}$. Compared to the RMS score in Eq. 4.2, the sign of each $z$-score value is kept. The target variable is then the SEC $z$-score,

$$y_n = \frac{x_{\text{SEC},n} - \mu_{\text{SEC}}}{\sigma_{\text{SEC}}}, \tag{4.11}$$

where $\mu_{\text{SEC}}$ and $\sigma_{\text{SEC}}$ are computed from the training data.

A linear model is then used to predict the $y_n$,

$$\hat{y}_n = \beta_0 + \sum_{f \in \mathcal{F}_{\text{reg}}} \beta_f z_{n,f}, \tag{4.12}$$

where $z_{n,f}$ is the signed $z$-score of feature $f$ for route $n$. Since the number of input features is relatively large in terms of regression, and several road features may be correlated, the coefficients are estimated using ridge regression,

$$(\hat{\beta}_0, \hat{\boldsymbol{\beta}}) = \arg\min_{\beta_0, \boldsymbol{\beta}} \sum_{n \in \mathcal{T}} \left( y_n - \beta_0 - \sum_{f \in F_{\text{reg}}} \beta_f z_{n,f} \right)^2 + \alpha \sum_{f \in F_{\text{reg}}} \beta_f^2, \tag{4.13}$$

where $\mathcal{T}$ is the training set, and $\alpha$ is the regularisation parameter. Missing feature $z$-scores are set to zero before fitting and prediction.

The predicted energy score is defined as the absolute value of the predicted signed SEC $z$-score,

$$\hat{S}_{\text{energy},n} = |\hat{y}_n|. \tag{4.14}$$

The predicted score has the same interpretation as the energy score in Eq. 4.9, meaning that a route with a low score is expected to have an SEC close to the customer average. The predicted energy score may either be used directly to rank routes, or be combined with the road feature-based representativeness score as

$$S_{\text{total,pred}} = S_{\text{total}} + w_{\text{pred}}\hat{S}_{\text{energy}}. \tag{4.15}$$

The predicted energy score is mainly used for experimental purposes to analyse how well energy can be estimated from the road features of the sOC.

# 5
# Results

This chapter presents the results obtained from applying the presented methodology to customer missions and generated routes. The customer-specific road feature distributions are analysed first, as they define the reference sOC. The route selection is presented, followed by a comparison between the representativeness score and the estimated energy consumption, and, lastly, energy score prediction.

## 5.1 Customer mission distance distribution

Figure 5.1 shows the distribution of total route distance across the customer transport missions. The route lengths vary across missions, with the longest route more than three times as long as the shortest. The mean distance is approximately 299 km, and the standard deviation is about 79 km. This variation is reported for context and discussed in Section 6.1, but total distance is not included as a feature in the representativeness scoring.

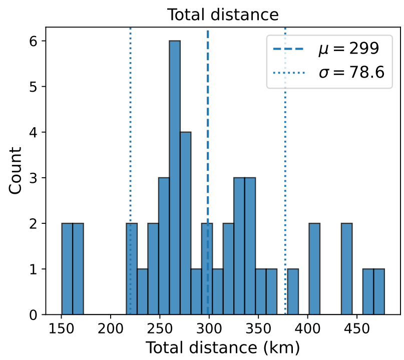



**Figure 5.1:** Distribution of total distance across a collection of a customer's transport missions.

## 5.2 Customer-specific road features

The statistical properties of the customer-specific sOC are visualised through histograms of the estimated features. These distributions form the basis for the representativeness score, as the mean and standard deviation of each feature are used in the $z$-score formulation (Eq. 4.1). The histograms also provide insight into the variability and typical ranges of the stochastic parameters for road characteristics across the collection of transport missions.

The figures in this section show the untransformed estimated feature distributions, while the log-transformed versions are shown in Appendix A. All histograms are shown with 30 bins to provide a consistent visual comparison between features and road types.

### 5.2.1 Road type ratio distributions

Figure 5.2 shows the distributions of road type ratios, the proportions of distance driven on each road type for the transport operations. Rural driving dominates customer operations, with a mean ratio of approximately 63.4%, while urban and highway ratios are lower, at 17.8% and 18.8%, respectively. The highway ratio also includes routes with little or no highway driving, indicating that highway driving is absent or negligible for some missions in the customer data.

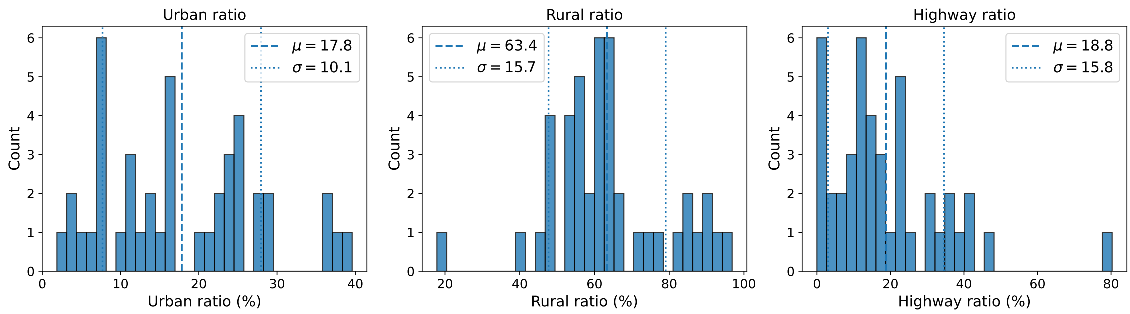



**Figure 5.2:** Distribution of road type ratios across a collection of a customer's transport operations.

### 5.2.2 Urban feature distributions

Figure 5.3 shows the distributions of the urban secondary stochastic parameters. The traffic control device features show a clear right skew in the lambda values, indicating that only some routes have significantly higher device densities than the others. The stop sign intensity is concentrated near zero, while the traffic light, yield sign, and pedestrian crossing intensities show greater variation across routes. The average speed is relatively concentrated around 43-50 km/h, with some left skewness. The slope grade mean is centred close to zero, while slope grade variance is right-skewed and average hill length is spread around 85 m. The curvature parameters show a wider spread, particularly the curvature radius variance, while two observations of the average curve length distort the histogram. Overall, the urban feature set contains both highly variable and strongly skewed parameters.

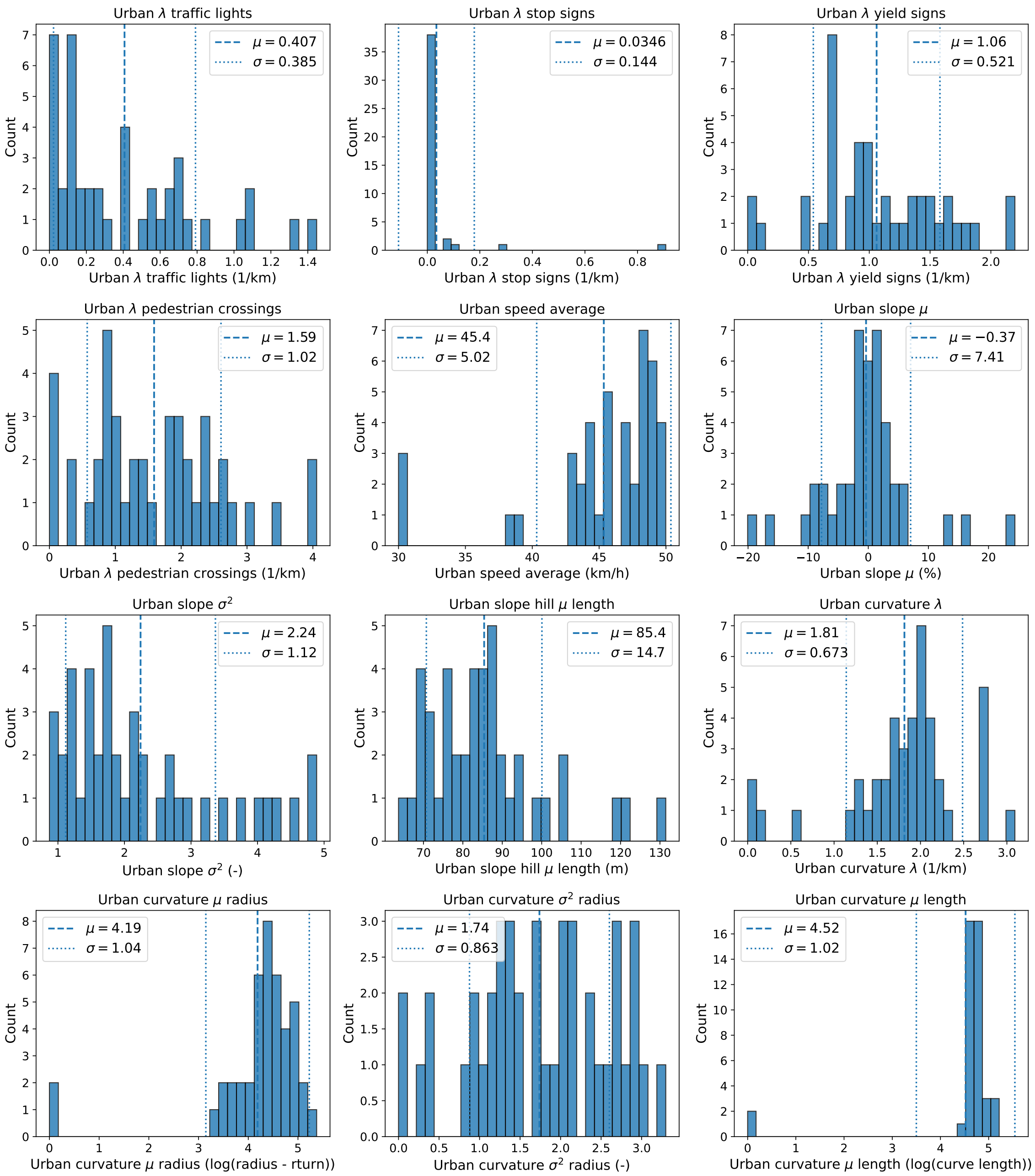



**Figure 5.3:** Distribution of secondary stochastic road parameters for urban road type across a collection of a customer's transport operations.

### 5.2.3 Rural feature distributions

Figure 5.4 shows the distributions of the rural secondary stochastic parameters. The traffic control device intensities for traffic lights, stop signs, and pedestrian crossings are right-skewed and concentrated near zero, while the yield sign intensity has a wider distribution. The average speed shows variation while peaking at around 72 km/h. The slope grade mean histogram peaks around zero with some variation in the interval between $-4\%$ to $3\%$. The slope variance is slightly right-skewed, while the average hill length peaks at around 115m with varying values. The curvature intensity and radius mean histograms are more concentrated than those for urban, while curvature radius variance and mean hill length show a peak with variability around it.

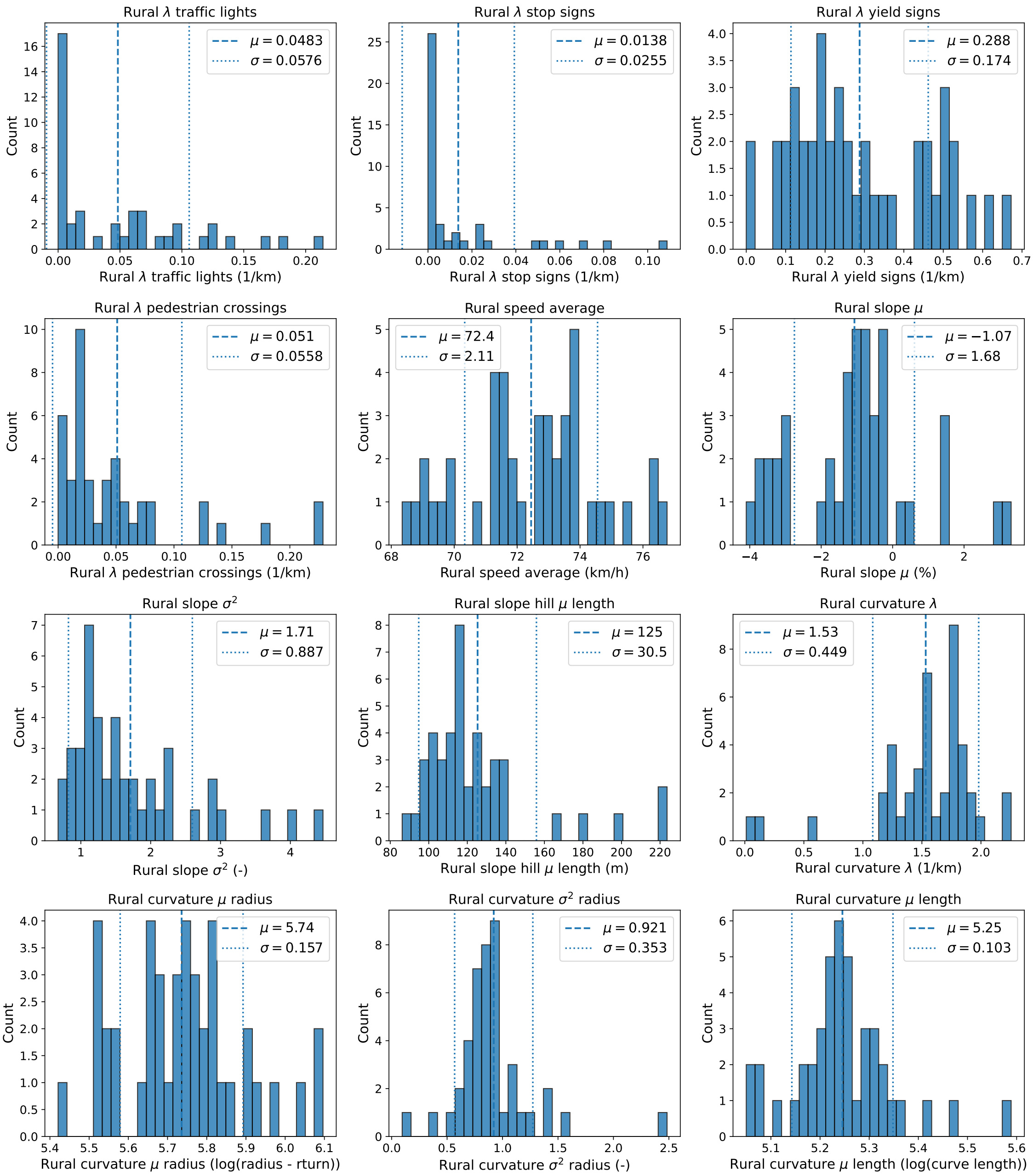



**Figure 5.4:** Distribution of secondary stochastic road parameters for rural road type across a collection of a customer's transport operations.

### 5.2.4 Highway feature distributions

Figure 5.5 shows the distributions of the secondary stochastic parameters for the highway. The traffic control device intensities are zero for almost all missions, indicating that such devices are absent on the observed highway segments. The average speed is constant at 90, as it is the only possible speed limit. The behaviour of the traffic control device features and the average speed are the reasons these variables are omitted from the highway scoring features, as discussed in Section 4.1.4. The slope grade mean shows greater variation than both previous road types, with most values between −5% and 10%. The slope grade variance is right-skewed with some variation, while the mean hill length has a relatively wide spread and longer hill lengths than previous road types. The curvature intensity varies, while the curvature radius variance histogram is distorted by four zero counts. The mean curvature radius and mean curve length distributions include a few isolated low values compared to most observations.

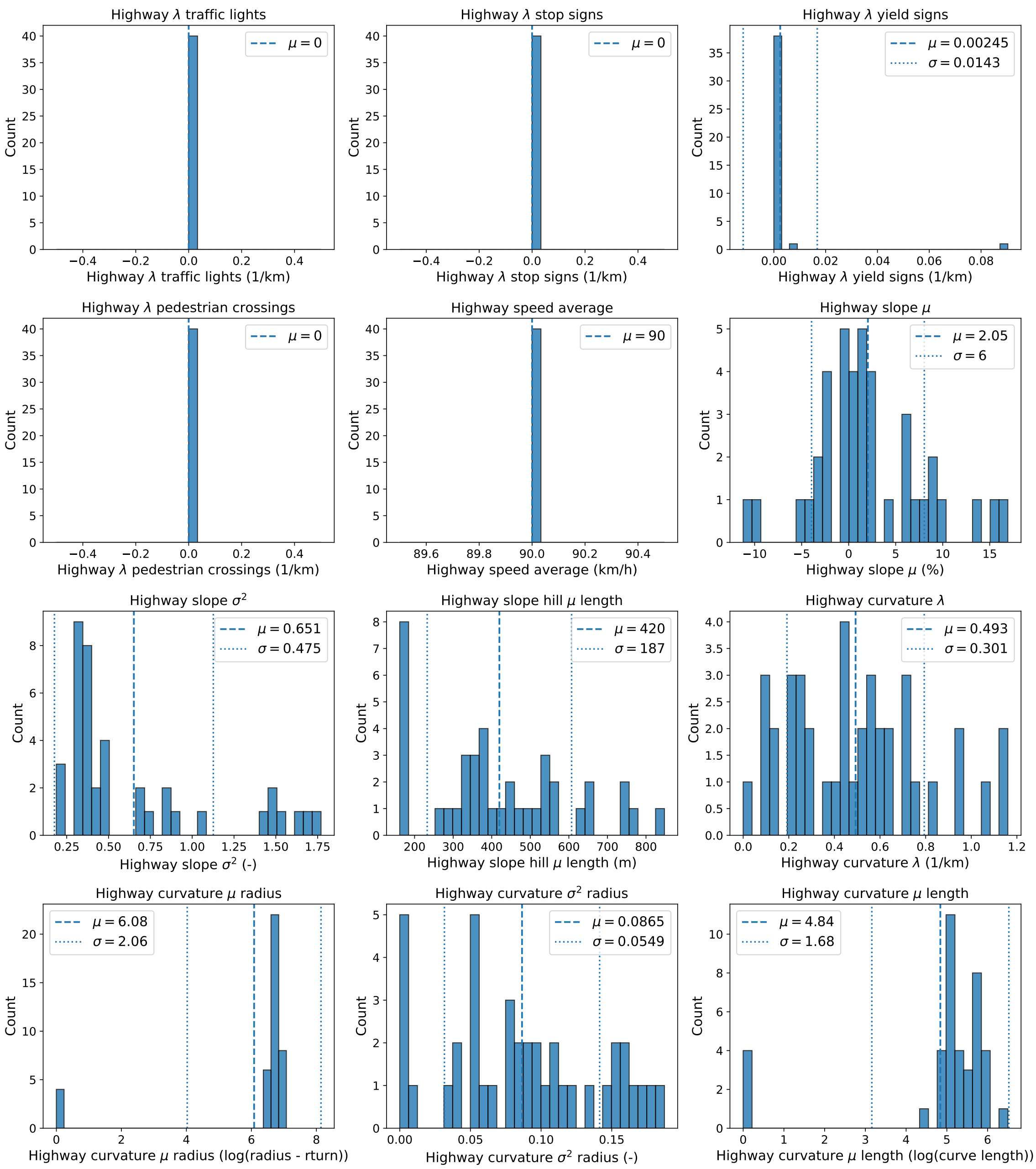



**Figure 5.5:** Distribution of secondary stochastic road parameters for highway road type across a collection of a customer's transport operations.

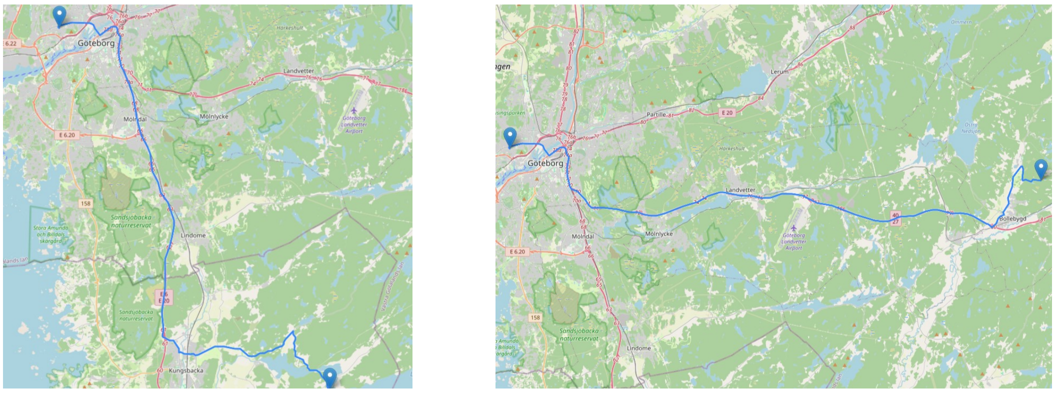
## 5.3 Test route selection

Using the isoline route generation as described in Section 3.1.1, a set of 57 routes was generated with Volvo Group Lundby as origin, or more precisely, the point (57.717107, 11.920933). Both the generated routes and the customer log files were evaluated using the representativeness score, and then ranked with the scores of the three best-scoring and three worst-scoring routes presented in Table 5.1.

**Table 5.1:** Best and worst route representativeness scores for customer log files and generated routes.

| Customer log files | Generated log files |
|---|---|
| 0.769 | 1.690 |
| 0.790 | 1.716 |
| 0.873 | 1.850 |
| 3.193 | 9.340 |
| 3.579 | 10.055 |
| 3.892 | 11.298 |

The best- and worst-scoring routes are shown in Figure 5.6, with maps generated by the script as described in Section 3.1.2.




**(a)** Best-scoring generated route. **(b)** Worst-scoring generated route.

**Figure 5.6:** Maps over the best- (a) and worst-scoring (b) generated routes.

## 5.4 Estimated energy consumption

Figure 5.7 shows the distribution of estimated SEC across the customer log files. The estimated SEC values have a mean of approximately 2.72 kWh/km and a standard deviation of approximately 0.203 kWh/km. The variation shown in the figure mainly reflects differences in the road data for the logged missions, as the energy-estimation script indicates, since weather variables and weight were kept constant, and mission stops were excluded.

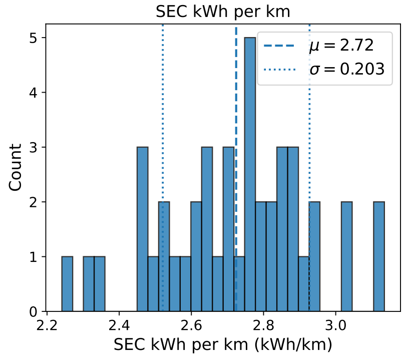
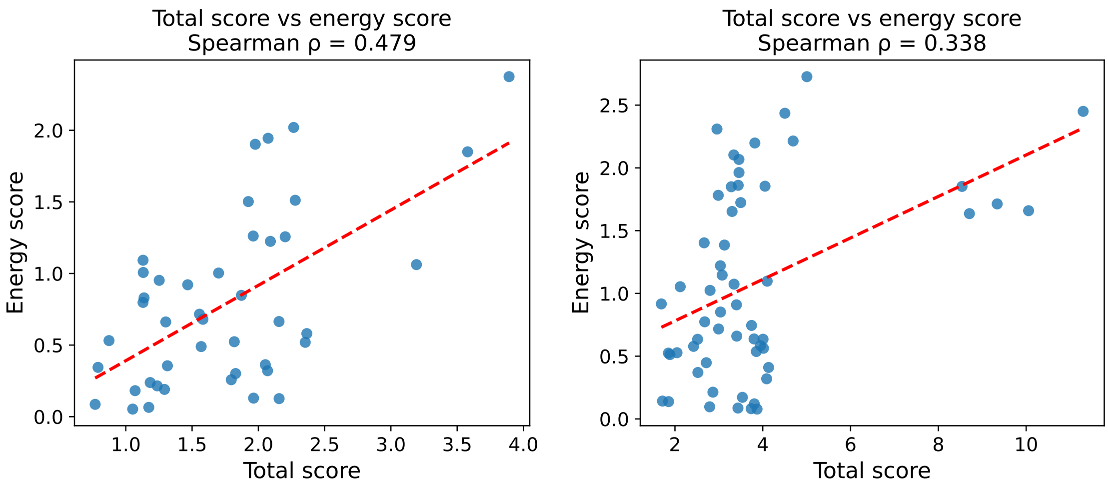



**Figure 5.7:** Distribution of estimated SEC across a collection of a customer's transport missions.

## 5.5 Representativeness score to energy

To investigate whether the representativeness score aligns with the estimated energy consumption, the representativeness score (total score) was plotted against the energy score in Figure 5.8. The representativeness score is used with weights as described in Section 4.1. Figure 5.8a contains customer log files, while Figure 5.8b contains generated routes. Each point represents one log file, and both scatter plots include best-fit lines. To obtain a quantitative measure of the strength of the correlation between the representativeness score and the energy score, the Spearman rank correlation was computed, yielding 0.479 for the customer log files and 0.338 for the generated routes.A Spearman correlation of 1 indicates a perfect positive monotonic relationship, -1 a perfect negative monotonic relationship, and 0 no monotonic rank relationship.




**(a)** Customer log files **(b)** Generated log files

**Figure 5.8:** Representativeness score plotted against energy score as scatter plots with best-fit lines. Each point represents one log file.

Since the importance of the road type ratio score and the road type-specific scores is a modelling choice, a sensitivity analysis was performed by varying the ratio score weight $w_{\text{ratio}}$ and the road type-specific weights, and computing the Spearman rank correlation between the representativeness score and the energy score for the different combinations.

Two sets of road type-specific weights were compared. In the first case, the road type scores were weighted equally, with $w_{\text{urban}} = w_{\text{rural}} = w_{\text{highway}} = 1/3$. In the second case, the original formulation was used, where the road type scores were weighted by the mean road type ratios in the customer data, $w_i = \mu_{\text{ratio},i}$. The results are shown in Table 5.2.

**Table 5.2:** Spearman rank correlation between representativeness score and energy score for different ratio and score weights. Two choices of road type score weights are compared: equal weights, $w_{\text{urban}} = w_{\text{rural}} = w_{\text{highway}} = 1/3$, and weights based on the mean road type ratios, $w_i = \mu_{\text{ratio},i}$.

| $w_{\text{ratio}}$ | Equal road type weights customer log files | Ratio weights customer log files | Ratio weights generated log files |
|---|---|---|---|
| 0.0 | 0.360 | 0.290 | **0.466** |
| 0.2 | 0.460 | 0.410 | 0.429 |
| 0.4 | 0.499 | 0.463 | 0.390 |
| 0.6 | **0.523** | 0.472 | 0.353 |
| 0.8 | 0.507 | **0.486** | 0.347 |
| 1.0 | 0.485 | 0.479 | 0.338 |
| 1.2 | 0.488 | 0.463 | 0.334 |
| 1.4 | 0.480 | 0.459 | 0.315 |
| 1.6 | 0.479 | 0.463 | 0.298 |

The highest correlation when using equal road type weights was obtained with $w_{\text{ratio}} = 0.6$, with a Spearman correlation of 0.523. When using mean road type ratios as weights, the highest correlation with customer log files was obtained with $w_{\text{ratio}} = 0.8$, with a Spearman correlation of 0.486. For mean road type ratio weights on generated routes, the highest correlation was observed with $w_{\text{ratio}} = 0.0$, yielding a Spearman correlation of 0.466. The result should be interpreted as an analysis of this specific data set and customer rather than as a generally optimal choice of weights.

## 5.6 Energy prediction

The energy prediction model described in Section 4.3 was evaluated using 5-fold grouped cross-validation on 106 route segments from the customer log files, each 100 km in length. Segments from the same original route were kept in the same fold, so that each route did not contribute to both the training and validation sets. The ridge regularisation parameter was set to $\alpha = 55.0$.

Two correlations were computed for each validation fold. The first was the Spearman correlation between the predicted signed SEC $z$-score $\hat{y}$ and the corresponding actual signed SEC $z$-score $y$. The second was the Spearman correlation between the predicted energy score $\hat{S}_{\text{energy}}$ and the actual energy score $S_{\text{energy}}$. The mean correlation over the five folds was 0.839 for the signed $z$-scores and 0.532 for the energy scores. The energy score correlations ranged from 0.173 to 0.681 over the five folds, while the signed correlations ranged from 0.758 to 0.916.

A final prediction model was taken as the mean regression feature coefficients across the folds and applied to the generated routes. Figure 5.9 shows the predicted energy score plotted against the energy score for the generated routes, with a Spearman correlation of 0.645.

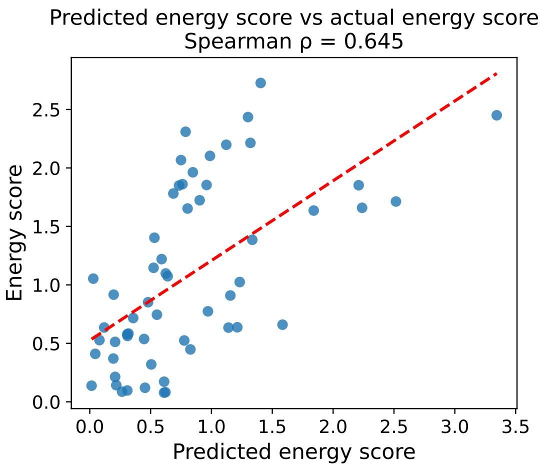



**Figure 5.9:** Predicted energy score plotted against energy score as a scatter plot with best-fit line. Each point represents one generated log file.

# 6
# Discussion

This chapter discusses the findings presented in the results and their meanings for selecting representative test routes. The discussion focuses on variation in mission length, score weighting, the relationship between road features and estimated energy consumption, and how these factors affect the interpretation of the results. Use of the score for selecting representative customer log files is also discussed.

## 6.1 Mission length variation

Figure 5.1 shows the variation in total route distance across the customer missions. The total distance is not included in the representativeness score, but the population statistics are computed from feature summaries per route. Consequently, each mission contributes one set of estimated parameters to the population, regardless of route length. This is a modelling choice to include as many logged missions from a single customer as possible. However, it also means that short missions influence population statistics in the same way as longer missions, even though estimates from short missions may be less reliable and more sensitive to outliers. Unusual road segments may yield extreme parameter values more often on shorter missions, while longer missions are more likely to yield estimates that better reflect overall usage.

This should be kept in mind when interpreting the results. Shorter missions are not uninformative, but all missions currently contribute equally to the scoring, even though their parameter estimates differ in reliability.

## 6.2 Score to energy correlation

The analysis in Table 5.2 shows that the correlation between representativeness scores and energy scores depends on the choice of score weights. Equal road type weights gave a slightly higher maximum correlation than using mean road type ratios as road type-specific weights. The differences are relatively small, both between the chosen score weights and the ratio weights near the optimal values. However, for the customer log files, an optimal ratio weight ($w_{\text{ratio}}$) appears to be in the range of 0.4 to 0.8 when using equal road type weights, and 0.6 to 1.0 when using road type ratios as weights. Surprisingly, given the first two results, the highest correlation was observed without any ratio score for the generated routes with equal road type weights.

The results suggest that the representativeness score is related to estimated energy consumption, even though it is based solely on road feature deviations. This is expected, since road type, topography, curvature, and speed characteristics are expected to influence energy consumption. However, the representativeness score is not specifically designed to optimally estimate energy consumption, but rather serves as a statistical model to compare deviations in energy-related road characteristics between candidate routes and a customer-specific road profile.

The results also suggest that the score weights could be optimised further with respect to energy consumption if the goal is to find a route representative of energy consumption. However, a more systematic analysis would be required to find weights that generalise well, including more log files, with greater variation and, preferably, a separate validation set. With the approach used, the weights can be optimised for one customer.

## 6.3 Energy score prediction

As seen in Section 5.6, the higher correlation for the signed $z$-scores indicates that the regression model captures whether a route is expected to have a higher or lower SEC than the customer average. The lower correlation for the absolute energy score indicates that the prediction is less precise when the sign is removed by taking the absolute value and comparing only the distance from the mean SEC.

Interestingly, when using the prediction model on the generated routes, the correlation of the actual energy score was higher than the mean correlation across the five folds, 0.645, compared to 0.532. The correlation was also higher than the largest correlation obtained between the representativeness score and energy score for generated routes in the weight analysis in Table 5.2.

This result suggests that the road features used in the representativeness score are relevant for estimating energy consumption. However, the regression model is fitted specifically to predict the estimated SEC, while the representativeness score is designed to compare routes statistically. The predicted energy score should therefore be interpreted as a complement to the representativeness score, rather than a replacement when the primary goal is to match the customer's road operating conditions, and the combination of both can be used as a final score as proposed in Eq. 4.15.

## 6.4 Use the score for selecting customer log files

The representativeness score can be used not only to evaluate generated candidate routes but also to rank the customer log files themselves. For a customer with a large number of logged missions, each log file can be compared against the customer-specific population in the same way as a generated route. Low-scoring log files can be interpreted as missions typical of the customer, while high-scoring log files can be interpreted as unusual or outlier missions.

This could be useful when selecting log files for further analysis, simulation, or validation. Instead of choosing log files at random, the score provides a systematic way to select missions that are representative of the customer's usual operations. At the same time, high-scoring log files may represent rare but relevant operating conditions. The score could therefore be seen as a tool for identifying customer-typical missions, or unusual missions. Lastly, the parameters included in the scoring can be adjusted to, for example, only include slope parameters in situations where specific road characteristics are of interest.

## 6.5 Conclusions

A method for selecting real-world test routes that are representative of a customer's road-related operating conditions has been developed. Customer log data, along with corresponding road attributes, were used to construct dOCs, estimate sOC parameters, and define a representativeness score based on deviations from customer feature distributions.

The results show that the score can be used to rank routes, both generated routes and customer log files. Routes with low scores are interpreted as being closer to the customer's typical road operating conditions, while routes with high scores represent unusual missions. Comparison with estimated energy consumption indicates that the representativeness score is related to energy consumption, although it is not designed to directly optimise for energy consumption.

Regression-based energy score prediction showed that road features are related to SEC, and therefore relevant for estimating a route's SEC. However, this score should be seen as a complement to the representativeness score, since the main purpose of the method is to match road characteristics rather than energy consumption.

The thesis shows that the OC can be used to describe customer usage statistically and to evaluate route representativeness. The proposed method enables comparison of routes against customer-specific road characteristics, thereby contributing to statistically representative route selection.

## 6.6 Future work

The route selection method could be improved to be more robust and useful for real-world applications. Some such future improvements are proposed: inclusion of additional OC categories, parameter estimation analysis, and a more consistent treatment of mission length variation.

### 6.6.1 Weather, traffic and mission conditions

A natural step would be to extend the method to include weather, traffic, and mission conditions based on the OC framework, rather than only road-related factors. In this thesis, these parts were omitted due to time constraints and complexity, even though they affect how a vehicle is operated in practice. At the same time, the OC framework is already built around the four categories: road, weather, traffic, and mission, so adding these parts would fit well with this work.

Including these categories would improve route selection by matching candidate routes to a larger set of operating conditions. Romano [14] showed that extending the OC with stochastic models for weather, traffic, and mission properties enables a more accurate description and simulation of transport missions. The framework is meant to be parsimonious and modular, so these additions do not have to be implemented all at once. Instead, they could be added one at a time.

If these new categories are added, the representativeness score would need to be extended as well. A more complete measure should therefore include features from these categories as well, so that the selected route better matches the customer's full operating conditions rather than only the road characteristics.

### 6.6.2 Parameter estimation

The amount of data required to obtain reliable estimates should be studied with respect to both mission length and distance within each road type, as well as the number of customer log files. Since the secondary features are estimated separately for urban, rural, and highway roads, a long mission may still give unreliable estimates within a road type that is only driven over a short distance. This is especially important for rare features, such as some traffic control devices, where a small amount of data can lead to unstable estimates.

Future work could compute feature statistics using varying amounts of data and compare estimates obtained from different data sets. This could be done by obtaining more customer log files to work with, truncating missions to different lengths, or requiring minimum distances within each road type. The same idea could also be used to study how many log files are needed before the estimated population distributions and route rankings become stable.

An analysis as described could lead to criteria for data before using the developed scripts, for example, for when a log file should be included, the minimum number of log files from a customer, or when a road type-specific estimate should be excluded.

### 6.6.3 Energy prediction

The regression model used for energy prediction performed well, however, the prediction accuracy could be improved by training on a larger and more differing data set in order to generalise better and be more robust. In the current implementation, only segments from log files of one customer were used as training data, which means that the training data is similar.

In this work, an energy estimation script was used and the regression was trained against the estimations obtained. Instead, actual measured energy consumption per kilometre could be used as ground truth, which would result in more accurate training data. The energy score could also be calculated from the measurements, so that the correlation between representativeness score and energy score could be re-calculated.

### 6.6.4 Mission length variation

In future work, the variation in mission distances could be addressed by introducing minimum-distance requirements, truncating log files to equal lengths, or weighting missions when computing the population mean and standard deviation based on the amount of data within each road type. This would preserve the interpretation of the estimated parameter, while reducing the influence of poorly supported estimates from short missions. The parameter values themselves should not be modified by distance weighting, as this would yield false parameter values that lose their meaning. Instead, weighting should be applied only when aggregating estimates across missions.

# Bibliography


[1] Dobrow, R. P. (2016). *Introduction to stochastic processes with R*. John Wiley & Sons.

[2] Edlund, S., & Fryk, P. O. (2004). The right truck for the job with global truck application descriptions. *SAE Technical Paper*, SAE Technical Paper 2004-01-2645. DOI: 10.4271/2004-01-2645.

[3] Emvin, C. (2023). *Assessing the feasibility of replacing a specific diesel truck with a battery electric vehicle using the operating cycle format*. Master's thesis, Chalmers University of Technology.

[4] Fontaras, G., Martin, R., Dilara, P., Hausberger, S., & Anagnostopoulos, K. (2013). The development of a simulation tool for monitoring heavy-duty vehicle CO2 emissions and fuel consumption in Europe. *SAE Technical Papers*, SAE Technical Paper 2013-24-0150. Society of Automotive Engineers (SAE International). DOI: 10.4271/2013-24-0150.

[5] HERE Technologies. *Flexible Polyline Encoding*. GitHub repository. Available at: `https://github.com/heremaps/flexible-polyline` (Accessed: 2026-02-02).

[6] HERE Technologies. *Introduction to HERE Isoline Routing v8*. HERE Isoline Routing API v8 Developer Guide. Available at: `https://docs.here.com/routing/docs/isoline-routing-intro` (Accessed: 2026-02-02).

[7] HERE Technologies. *Introduction to HERE Route Matching API v8*. HERE Route Matching API v8 Developer Guide. Available at: `https://docs.here.com/routing/docs/intro-route-matching` (Accessed: 2026-02-02).

[8] HERE Technologies. *Introduction to HERE Routing API v8*. HERE Routing API v8 Developer Guide. Available at: `https://docs.here.com/routing/docs/routing-intro` (Accessed: 2026-02-02).

[9] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics, Vol. 103. Springer. DOI: 10.1007/978-1-4614-7138-7.

[10] Lang, A. & Petersson, A. (2016). *Financial Time Series (TMS087) lecture notes*. Available at: `https://utb.math.chalmers.se/Stat/Grundutb/CTH/tms087/1516/lecture-notes/FTS-lecture-notes.pdf`

[11] Nyberg, P. (2015). *Evaluation, Generation, and Transformation of Driving Cycles*. PhD thesis, Linköping University.

[12] OpenStreetMap contributors (n.d.). *About OpenStreetMap*. Available at: `https://www.openstreetmap.org/about` (Accessed: 2026-02-02).

[13] Pettersson, P. (2019). *Operating cycle representations for road vehicles*. PhD thesis, Chalmers University of Technology.

[14] Romano, L. (2023). *The operating cycle representation of road transport missions*. PhD thesis, Chalmers University of Technology.

[15] Romano, L., Emvin, C., Bruzelius, F., Johannesson, P., Andersson, R., & Jacobson, B. (2023). Stochastic modeling of mission stops and variable cargo weight for heavy-duty trucks. In *Proceedings of the 2023 IEEE Vehicle Power and Propulsion Conference (VPPC)* (pp. 1–8). DOI: 10.1109/VPPC60535.2023.10403336.

[16] Romano, L., Godio, M., Johannesson, P., Bruzelius, F., Ghandriz, T., & Jacobson, B. (2023). Development of the Västra Götaland operating cycle for long-haul heavy-duty vehicles. *IEEE Access*, 11, 73268–73302. DOI: 10.1109/ACCESS.2023.3295989.

[17] Svensson, J., & Bertilson, H. (2025). *Stochastically modelling road topography*. Master's thesis, Chalmers University of Technology.

[18] Volvo Group (2025). *Volvo Group Annual Report 2024*. Available at: `https://www.volvogroup.com/content/dam/volvo-group/markets/master/events/2025/annual-reports/volvo-group-annual-report-2024.pdf` (Accessed: 2026-02-02).

# A

# Appendix A

## A.1 Log-transformed feature distributions

This appendix shows the feature distributions after the logarithmic transformations described in Section 4.1.2. The figures are included for visual comparison to the feature distributions in Section 5.2. The plots include only the features that have been log-transformed. The effect of the transformation depends on the underlying distributions, and the transformation does not make all features approximately normal, especially not for variables with many zero observations.

### A.1.1 Urban log-transformed feature distributions

Figure A.1 shows the distributions of urban log-transformed features. Traffic light lambda is more spread out and less right-skewed, but it still shows a high bar at the beginning. For the stop sign lambda, the transformation does not visually change much, with only a small change noticeable, a few bars spread out more. Pedestrian crossing lambda is more symmetric. The slope variance tends more toward a normal distribution, as does the average hill length distribution.

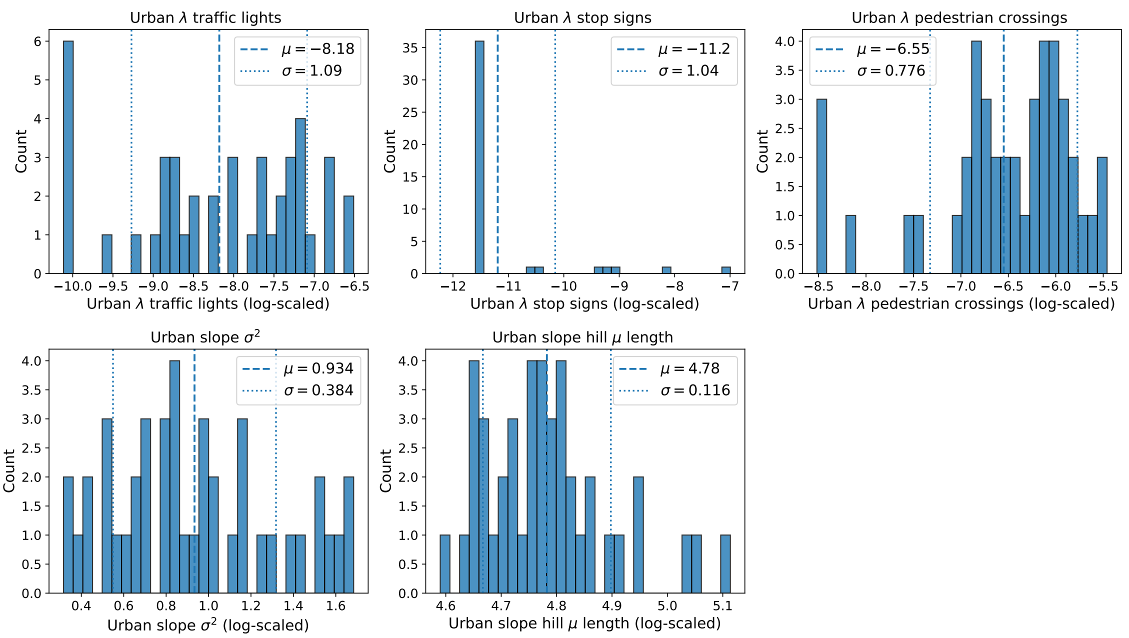



**Figure A.1:** Urban feature distributions after the log transformations.

### A.1.2 Rural log-transformed feature distributions

Figure A.2 shows the distributions of rural log-transformed features. In this case, the transformation does not change the distributions of the traffic light and stop sign lambdas much. The yield sign and pedestrian crossing lambda distributions and slope grade variance are more symmetric post-transformation, as the right tails are compressed.

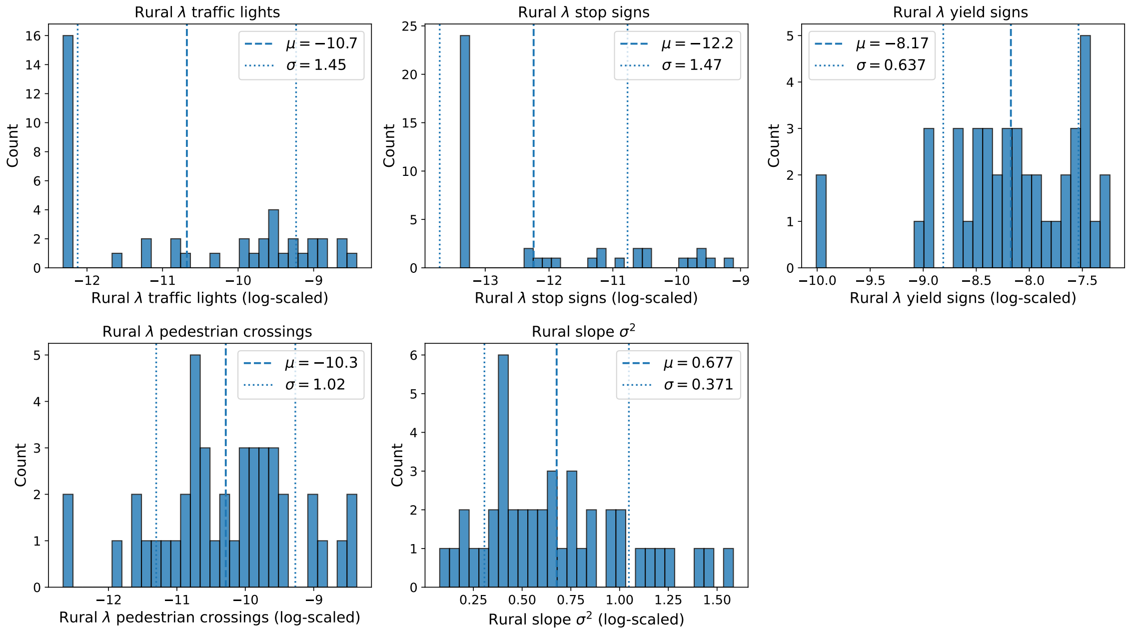



**Figure A.2:** Rural feature distributions after the log transformations.

### A.1.3 Highway log-transformed feature distributions

Figure A.3 shows the distributions of highway log-transformed features. Both the slope grade variance and average hill length histograms show greater symmetry after transformation.

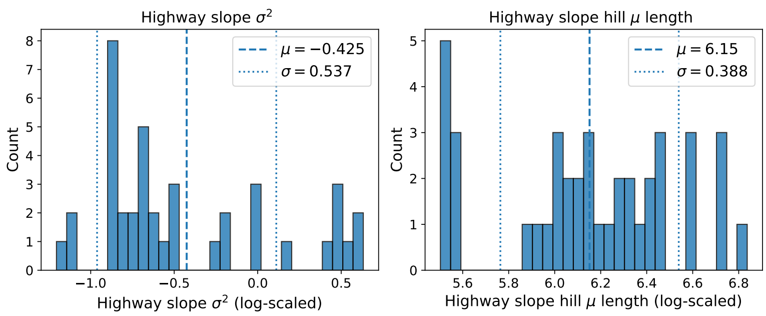



**Figure A.3:** Highway feature distributions after the log transformations.

**DEPARTMENT OF MECHANICS AND MARITIME SCIENCES**
**CHALMERS UNIVERSITY OF TECHNOLOGY**
Gothenburg, Sweden 2026
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY