Stocking up for winter

Quantifying the trade-off between robust and efficient fleet scheduling for public transit operating under energy consumption uncertainty

MSc Thesis

S.F.J. Buisman



Delft University of Technology

EBS

TU Delft

Stocking up for winter

Quantifying the trade-off between robust and efficient fleet scheduling for public transit operating under energy consumption uncertainty

by

S.F.J. Buisman

to obtain the degree of Master of Science in Transport, Infrastructure and Logistics

at the Delft University of Technology.

Student number:	5957915	
Project duration:	December 2025 - October 2026	
Thesis committee:	N. van Oort ,	TU Delft, chair
	P.S.A. Stokkink,	TU Delft, supervisor
	M. Meijers,	EBS Public Transportation, advisor
	J.E.B. Bickel,	EBS Public Transportation, advisor

Cover:	BYD K9UB electric bus, photo from www.busposities.nl made by user Sarah1763
Style:	TU Delft Report Style, with modifications by Daan Zwaneveld

Preface

After a few delays and unexpected detours, I have arrived at my final destination as a student. Before arriving at the TU Delft, I did a Bachelor's in landscape architecture. While I still hold great interest in the urban planning and design, I want to focus more on how people move from one place to another. In Wageningen I learned that good public transport is a requirement to create more liveable places and the courses in Delft helped me understand how to create good public transport. It was a recurring theme throughout my years as a Master student; even becoming a part-time bus driver in the process.

Not only as a driver, but also as a frequent user of public transport, I experienced a fair share of buses that did not show up. At the same time I am impressed by the speed in which the Dutch public transport sector has shifted towards zero-emission vehicles. This study has allowed me to combine my interest in reliable public transport with technicalities of the energy transition and hopefully deliver a contribution towards the improvement of both.

This research was not possible without my advisors at EBS, Julia Bickel and Maxim Meijers. Julia provided valuable insights on how EBS handles data and helped me to get going quickly by gathering large datasets and showing me around. Maxim shared his broad understanding of how public transport comes together and had good advice on how to keep the research connected to real operations. Further thanks go out to Julius and Renate, who created additional schedules, and all the colleagues who contributed their ideas on the matter and made the place not feel like work when talking about work.

I am grateful to my supervisors from the TU Delft, Patrick Stokkink and Niels van Oort. They gave practical tips in every step and took time to explain their suggestions. Niels shared his expertise on public transit and challenged me to think further than the numbers. Patrick proved to be invaluable in refining the methodology, which has led to a better framework.

Last, I want to thank my friends and family for their ongoing support throughout this journey. This is especially true for my girlfriend, Ceel, who encouraged me throughout the whole process. She patiently put up with my bus stories and my absence on thesis-heavy weekends. While among many friends it has become somewhat of a running joke whenever I mention the word 'bus', I am sure they share the idea that good public transport makes the difference between a place you live in and one you merely pass through.

S.F.J. Buisman
Delft, November 2026

Management summary

Context

Public transport operators in the Netherlands are moving to zero-emission buses. The energy consumption of a battery electric bus is uncertain: it depends strongly on, for example, the outside temperature. When a vehicle schedule is made, the operator has to assume a value for the energy consumption. This value determines how much energy redundancy the schedule contains, which means how much energy remains available on top of the expected consumption. If the assumed value is too low, buses run out of energy on cold days, trips are cancelled and passengers are left waiting. If it is too high, extra buses and kilometres are planned, which makes the schedule more expensive. This is the trade-off between a robust and an efficient schedule. Although the trade-off is intuitive, its size has hardly been quantified for the vehicle schedules of electric buses. Operators currently have to find the balance based on experience and rules of thumb, whereas the choice is made again for every new concession and timetable. This thesis aims to provide numbers to support that choice.

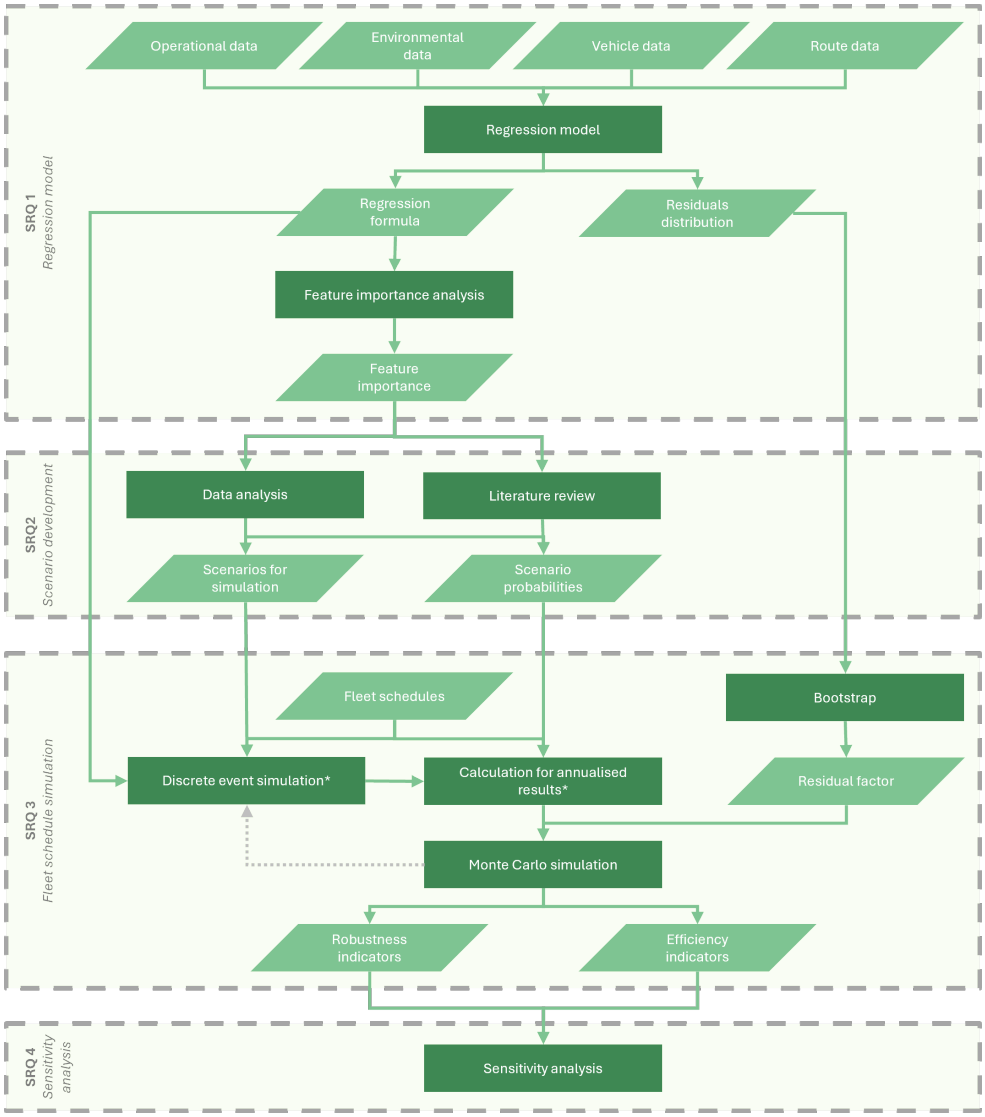


Figure 1: Schematic overview of the research approach

Objective and approach

The research quantifies how the energy consumption value that is assumed in the scheduling of electric buses shapes the trade-off between robustness and efficiency. It was carried out in cooperation with EBS Public Transportation, with data from the IJssel-Vecht concession and the vehicle schedules of the depots in Apeldoorn and Lelystad (the Netherlands). The approach consists of four steps, which are shown in figure 1. First, a data-driven model was built that estimates the energy consumption of a trip from nine factors, using about 700,000 trips from 2025. Second, scenarios were defined for the conditions that matter most: the temperature and the number of passengers. Third, a simulation of the schedule was used to assess how many trips are cancelled under these scenarios by the schedule that EBS actually operated in 2025, including the uncertainty that the model cannot explain. Fourth, four alternative schedules, generated with the planning software of EBS, were compared with this actual schedule. These schedules were planned with an energy consumption that is 10% or 20% lower or higher. A lower value means that the bus is assumed to have a longer range, and thus less energy redundancy. A higher value means a shorter assumed range and more redundancy. The schedules are therefore named after their redundancy in the figures: less (20% lower value), slightly less (10% lower), slightly more (10% higher) and more (20% higher) energy redundancy. The outcomes were compared in terms of completed trips and cost for passengers and for the operator.

Main findings

Outside temperature dominates the energy consumption. The energy consumption of an electric bus is lowest at about 19°C and 35% to 48% higher at the cold end of the temperature range in the data (around 0°C). The charge of the battery at the start of a trip and the number of passengers follow, but at a large distance. The model explains 76% of the variance in the energy consumption of a bus over a complete day, with 90% of the estimates within about 0.1 kWh/km (roughly 10%) of the measured value. For single trips it is much less accurate, because the battery charge is registered in whole percentages.

The actual 2025 schedules are robust down to about 5°C, but not at lower temperatures. At 5°C almost all trips are completed. At 0°C the share of completed trips drops to 99.4% in Apeldoorn and 96.4% in Lelystad, and at -5°C to 96.3% and 89.7%. Although such extreme days are infrequent (temperatures of -5°C are expected on less than 2% of the days), they cause a large part of the delay for passengers: about 63% in Apeldoorn and 38% in Lelystad.

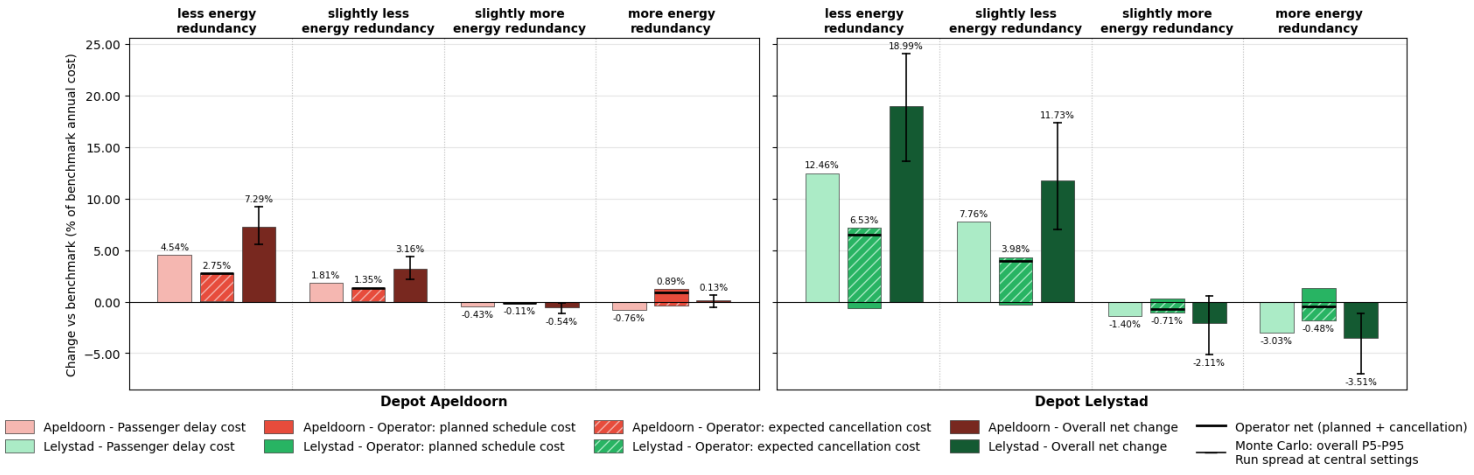


Figure 2: Change in cost for passengers and the operator, and overall net change in cost, of the alternative schedules compared with the actual schedule, without the cost of drivers. The whiskers show the 5th to 95th percentile of the simulation runs

Cutting energy redundancy to improve efficiency leads to higher cost. Figure 2 shows how the alternative schedules perform compared with the actual schedule. For every schedule, the first bar is the change in the cost of delays for passengers, the second bar the change in cost for the operator and the third bar the overall change. All values are percentages of the annual cost of the actual schedule for the operator, and a bar below zero is a saving. The schedules with less energy redundancy hardly

save on the planned schedule, but they lead to many more cancelled trips. Their overall cost is 3.2% and 7.3% higher in Apeldoorn and 11.7% and 19.0% higher in Lelystad. There is a limit to how far efficiency can be pushed, and if that limit is crossed, a higher efficiency even leads to higher cost.

The relationship is asymmetric: too little redundancy costs more than too much. A schedule with slightly more energy redundancy (10% higher value) lowers the overall cost in both depots (0.5% in Apeldoorn and 2.1% in Lelystad), whereas the schedule with more redundancy (20% higher value) pays off in Lelystad only. In Apeldoorn it needs an extra bus, which costs more than it saves. Raising the redundancy improves the share of completed trips only modestly, as it is already close to 100%.

The charging infrastructure matters. Apeldoorn, with fast chargers at the hub, is much less sensitive to the energy consumption value that is assumed in planning than Lelystad, where slower chargers are located at the depot, six kilometres away from the service area. Each extra charging trip, planned or unplanned, costs much more time in Lelystad. This is reflected in more extreme values for all cost indicators. Fast chargers at the hub also help in recovery during operations, as a bus that runs low on energy can be recharged within the service and is back sooner.

Driver schedules should not be ignored. More charging, driving to the charger and waiting also make the schedules of the drivers less efficient. Drivers are the largest cost component of an operator, so an estimate of this effect was added (figure 3). With this estimate, the benefit of slightly more energy redundancy remains in Apeldoorn (-0.4%), but disappears in Lelystad ($+1.6\%$), while the additional cost of less redundancy remains. The grey bars show how the outcome changes with the assumptions on the fine for a cancelled trip and the extent in which driver schedules efficiency is affected. The finding that less redundancy is costly holds for all combinations that were tested, but the finding that more redundancy pays off does not.

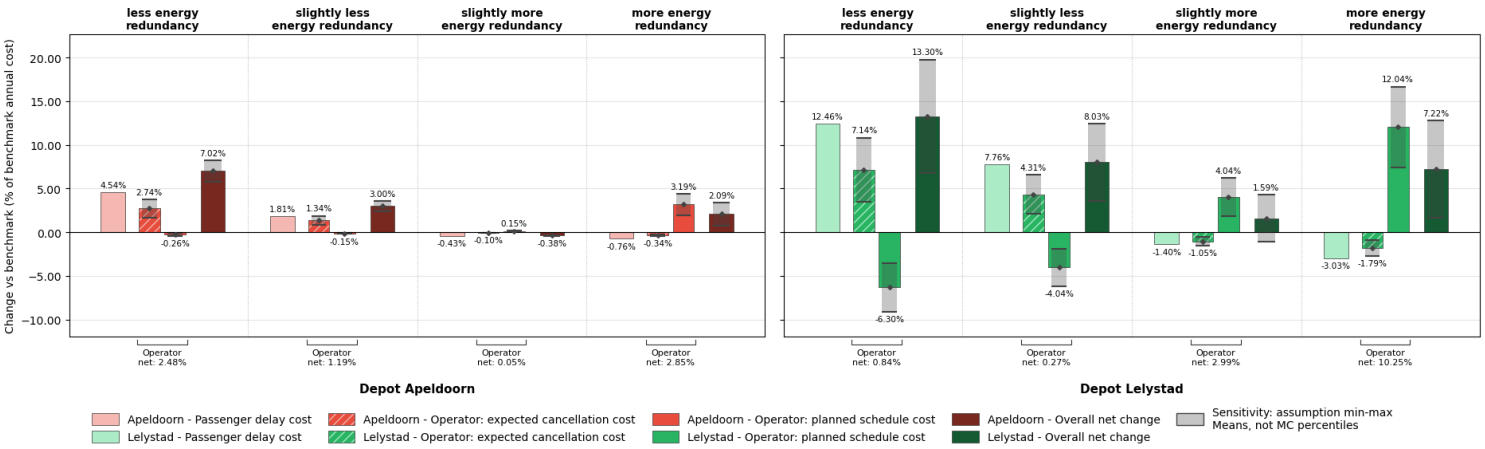


Figure 3: Change in cost compared with the actual schedule, including the estimated cost of drivers. The grey bars show the range of the overall change for different assumptions on the fine for a cancelled trip and the value of waiting time

Limitations

The results are the outcomes of two depots and an average climate year. The direction of the effects is likely to hold elsewhere, but the optimum and the amounts in euros depend on the climate, the schedule and the cost assumptions. The alternative schedules were generated with planning software that is a black box in this research. The fine for a cancelled trip and the cost of drivers are simplified, and the simulation only includes simple recovery measures, whereas in practice a driver or the control room can coordinate a bus swap before a bus has to cancel trips. The long-term effects of unreliable service, such as passengers who stop travelling by bus, are not included, which is an additional argument for leaning towards more robustness. Finally, the study focussed on low temperatures, whereas warm regions could face problems at high temperatures due to air conditioning.

Conclusion and implications

The findings follow from three scientific contributions. First, an energy consumption model was developed on a full year of real trips and validated on the level of a complete day of a bus, whereas the

reviewed studies report the accuracy per trip. It shows that a model that explains only 27% of the variance per trip can still be accurate enough (76% per bus per day) to assess a schedule, as the errors of individual trips partly cancel each other out. Second, a simulation framework was introduced that combines this model with scenarios and the unexplained variance, so that the robustness of a whole electric bus schedule can be tested for conditions that are rare but impactful. Third, the trade-off between robustness and efficiency was quantified for the energy consumption value that is used in planning, for passengers and the operator. This includes a test of how the conclusions depend on the cost assumptions.

Table 1 translates the case study into general findings for other concessions and operators. They also dive into the mechanisms behind the conclusions and explain under which conditions they hold.

Table 1: General findings derived from the case study

#	Conclusion	Mechanism and conditions
1	The climate of a concession determines when a schedule is at risk.	Applies to regions with a similar climate. In warmer regions, cooling at high temperatures may become the risk, which the data could not confirm.
2	Too little energy redundancy costs more than it saves, and the effect is asymmetric.	Cancellations carry cost for passengers and operator, while a completion rate near 100% leaves little to gain. Applies where cold weather raises the EC considerably.
3	Fast charging at the hub dampens the effect of redundancy, both in planning and in recovery during operations.	In planning, energy is restored quickly in a layover that is part of the service, without dead-heading. In operations, a bus that runs low can be recharged at the hub within the service and is back in service sooner, which limits the cancellations and shortens the recovery. Confounded by charger power, limited by the capacity of the chargers.
4	The value of redundancy depends on the ratio between cancellation cost and the efficiency loss of redundancy.	Redundancy pays off where fines and missed income are high and extra charging or vehicles are cheap.
5	Crew scheduling can decide whether robust fleet schedules pay off.	More charging, dead-heading and idle time make driver schedules less efficient. Operators may be hesitant to improve robustness without knowing the effect on crew schedules.
6	Rare extreme days dominate the cost of disturbances.	Cancelled trips on busy, cold days affect many passengers at once. Applies to every concession where one condition drives the energy consumption, but the extremes have to be weighted with their probability.

Two further insights are worth mentioning. A larger battery capacity could make the problem less relevant in the future, as a bus that can drive a complete day on a single charge does not depend on charging during the day. Furthermore, a schedule with less redundancy during warmer months is could save cost during that period, without harming robustness. It is not studied what the temperature threshold should be and for what period it should hold. That is a possible search direction for an operator, with the drawbacks of extra planning work and a change of driver duties twice a year.

Recommendations

EBS is recommended to determine the desired energy redundancy per depot, instead of applying one rule of thumb, as its value depends on the local context: the climate and the probability of extreme days, the location, power and capacity of the chargers, the cost of a cancelled trip and the effect on the driver schedules. This study warns against lowering the value, especially because of the effects on passengers. The framework can be used at other depots, as a stress test before a concession is operated (with similar material), and allows planning staff to explain to internal stakeholders how choices on robustness are made. Improving the data on energy consumption, in particular the resolution of the battery charge, should be a priority for the operator. Further research is recommended into a dedicated optimisation model for the schedules, testing the framework on more depots and regions, including crew scheduling, variation between years, and a richer energy consumption model with the road gradient, the number of stops and occupancy rates instead of check-ins.

Contents

Preface i

Management summary ii

Nomenclature xii

1 Introduction 1
- 1.1 Scope . . . 1
- 1.2 Research objective and contributions . . . 3
- 1.3 Outline . . . 4

2 Literature review on state-of-the-art 5
- 2.1 Energy consumption estimation . . . 5
  - 2.1.1 Modelling techniques . . . 5
  - 2.1.2 Influential parameters . . . 6
- 2.2 Robustness in public transit . . . 7
  - 2.2.1 Network robustness . . . 8
  - 2.2.2 Optimizing under uncertainty . . . 8
- 2.3 Fleet scheduling and charging optimization . . . 10
- 2.4 Conclusions from the reviewed literature . . . 11

3 Methodology 13
- 3.1 Regression model for energy consumption estimation . . . 15
  - 3.1.1 Speed dynamics indicator . . . 16
  - 3.1.2 Validation of model performance . . . 16
- 3.2 Scenario development . . . 17
- 3.3 Simulation of the fleet schedule . . . 17
  - 3.3.1 Discrete event simulation . . . 17
  - 3.3.2 Monte Carlo simulation . . . 19
  - 3.3.3 Indicators . . . 19
- 3.4 Sensitivity analysis . . . 22
  - 3.4.1 Fleet schedule optimization . . . 23
  - 3.4.2 Quantification of the trade-off . . . 23
  - 3.4.3 Additional analysis on the impact of crew scheduling . . . 24
- 3.5 Conclusions on methodology . . . 24

4 Case study 25
- 4.1 The IJssel-Vecht concession . . . 25
- 4.2 Selected depots . . . 25
  - 4.2.1 Network, fleet and infrastructure . . . 26
- 4.3 Benchmark and alternative schedules . . . 27
- 4.4 Cost parameters . . . 28

5 Data analysis 30
- 5.1 Data processing . . . 30
- 5.2 Data quality . . . 31
- 5.3 Factor value distributions . . . 32
- 5.4 Factor correlations . . . 33
- 5.5 Conclusions from the data analysis . . . 34

6 Results of the case study 35
- 6.1 EC model . . . 35

6.1.1 Factor relationships and importance . . . 35
6.1.2 Factor importance . . . 39
6.1.3 Validation on trip level . . . 41
6.1.4 Validation on vehicle task level . . . 44
6.1.5 Reflection on the EC model . . . 45
6.1.6 Conclusion on the EC model . . . 46
6.2 Scenario setup . . . 47
6.2.1 Future weather statistics . . . 47
6.2.2 Pax generation . . . 48
6.2.3 Omissions and simplifications . . . 50
6.2.4 Effect of simplifications . . . 50
6.2.5 Set of scenarios . . . 51
6.2.6 Reflection on the scenarios . . . 52
6.3 Simulation of the benchmark schedule . . . 53
6.3.1 Individual scenario performance for Lelystad . . . 53
6.3.2 Benchmark schedule performance . . . 56
6.3.3 Reflection on the benchmark results . . . 57
6.4 Sensitivity analysis . . . 59
6.4.1 Alternative schedules and correction of the benchmark . . . 59
6.4.2 Overall comparison without driver cost . . . 59
6.4.3 Robustness assessment . . . 61
6.4.4 Efficiency assessment . . . 62
6.4.5 Cost assessment . . . 64
6.4.6 Including driver cost . . . 64
6.4.7 Sensitivity to the cost assumptions . . . 66
6.4.8 Seasonal schedules . . . 66
6.4.9 Reflection on the trade-off . . . 67

7 Conclusion **69**
7.1 Answers to the research questions . . . 69
7.2 Scientific contributions . . . 70
7.3 General findings . . . 71
7.4 Recommendations for practice . . . 72
7.5 Limitations and recommendations for further research . . . 72

References **73**

A Vehicle tasks and bus changes **77**

B Speed dynamics as additional factor **79**
B.1 Speed dynamics indicators . . . 79
B.2 Data preparation . . . 80
B.3 Effect on the EC estimation . . . 81
B.4 Conclusion . . . 82

C Results without correction of the benchmark **83**

D AI statement **84**

List of Figures

1 Schematic overview of the research approach . . . ii
2 Change in cost for passengers and the operator, and overall net change in cost, of the alternative schedules compared with the actual schedule, without the cost of drivers. The whiskers show the 5th to 95th percentile of the simulation runs . . . iii
3 Change in cost compared with the actual schedule, including the estimated cost of drivers. The grey bars show the range of the overall change for different assumptions on the fine for a cancelled trip and the value of waiting time . . . iv

1.1 Electric bus planning process, derived from Perumal et al. (2022), with the steps that fall into the scope and the relationships of interest indicated by a dark green colour . . . 2

3.1 Schematic overview of the research approach . . . 14
3.2 Flowchart of the simulation model used to assess robustness . . . 18
3.3 Overview on the indicators on the four distinct levels . . . 20

4.1 Line networks and charging locations of the depots in Apeldoorn (left) and Lelystad (right). The hub is indicated by a network symbol and the chargers by a plug symbol . . . 26

5.1 Measured EC per kilometre plotted against trip distance. The grey lines indicate the EC that corresponds to one up to six discrete energy steps . . . 32
5.2 Distribution of values within every factor for both bus types . . . 33
5.3 Pearson correlation coefficient for each combination of two factors . . . 34

6.1 Partial dependence plots for all nine factors included in the model . . . 37
6.2 Drop in held-out R^2 when the values of a factor are shuffled (light bars) or when the factor is left out of the model (dark bars), for the 12m bus (blue) and 13m bus (orange) . . . 39
6.3 Example of the interaction effect between average speed and temperature, for the 12m bus (left) and 13m bus (right) . . . 41
6.4 Scatter plot (left) and histogram (right) with the trip level residuals . . . 42
6.5 Effect of trips with a distance below 10 kilometres on the residuals. The dotted lines indicate the 5^{th} and 95^{th} percentile of the residuals of all trips . . . 42
6.6 Trip level residuals plotted across the range of values of each factor . . . 43
6.7 Scatter plot (left) and histogram (right) with the vehicle task level residuals . . . 44
6.8 Normalised histogram of the daily mean temperature for the expected climate of 2027, based on the period 1991–2020 and the KNMI climate scenario for 2050 . . . 48
6.9 Distribution of the estimated daily total pax for the Apeldoorn (left) and Lelystad (right) depot. The dotted lines indicate the 5^{th} and 95^{th} percentile of cold school days, the solid line the median . . . 49
6.10 Estimated daily total pax against the daily mean temperature for Apeldoorn (left) and Lelystad (right) . . . 49
6.11 Residuals for vehicle tasks from the Apeldoorn and Lelystad depot calculated for both the base EC model and the EC model with simplifications for the simulation . . . 51
6.12 Overview of trip completion for all Lelystad vehicle tasks under the scenario with average pax and a day temperature of 0°C . . . 54
6.13 Overview of trip completion for all Lelystad vehicle tasks under the scenario with average pax and a day temperature of -5°C . . . 55
6.14 Net change in cost of the alternative schedules compared with the benchmark schedule, as a percentage of the annual operator cost of the benchmark schedule, without driver cost. The planned schedule cost is compared with the average of the -10 and $+10$ schedules . . . 60

6.15 Net change in cost compared with the benchmark schedule with the estimated driver cost included in the planned schedule cost, as a percentage of the annual operator cost of the benchmark schedule without driver cost. The grey bars show the range of the overall net change for different fines and values of idle hours 65

A.1 Bus changes per vehicle task and day for the Apeldoorn depot (blank = vehicle task not found in the data) 77
A.2 Bus changes per vehicle task and day for the Lelystad depot (blank = vehicle task not found in the data) 78

B.1 Speed profiles of trips with the most extreme global and line-specific scores 80
B.2 Filter results and distribution of the global score for the speed dynamics indicator 81
B.3 Comparison between the ridge regression base model without the speed dynamics score (left), and the model with the score (right) 82
B.4 Importance of the features in the ridge regression model with the speed dynamics score 82

List of Tables

1 General findings derived from the case study v

2.1 Literature search results 5
2.2 Overview of techniques and parameters used in the literature 7
2.3 Literature overview of robustness perspectives, subjects, and indicators 9
2.4 Comparison of problem characteristics and solution approaches 10
2.5 Comparison of operational characteristics 11
2.6 Comparison of the studies that come closest to quantifying the trade-off between robustness and efficiency 12

3.1 Data used in the EC analysis 15

4.1 Overview of the network, fleet and infrastructure of the two depots in the case study . . 27
4.2 Overview of the benchmark and alternative schedules. The range is the usable battery energy divided by the EC per kilometre, so a lower EC input implies a longer assumed range 27
4.3 Cost parameters used in the case study 28

5.1 Number of trips remaining after every step in the data processing 31

6.1 Estimated coefficients of the linear (β_j) and squared (β_{jj}) terms of the EC model, in kWh/km per standard deviation of the factor. The intercept β_0 is 0.8596 for the 12m bus and 0.9140 for the 13m bus. The 36 interaction terms are not shown 36
6.2 12m bus – PDP range within 5th–95th percentile range 36
6.3 13m bus – PDP range within 5th–95th percentile range 36
6.4 Top five interaction effects with the largest gain in held-out R^2 (compared to a model with only main and squared terms) 40
6.5 Model accuracy metrics on trip level 42
6.6 Model accuracy metrics on vehicle task level 45
6.7 Overview of the scenarios. Every temperature is combined with every pax level, which results in nine scenarios per depot. The expected number of days per year of a scenario (D_s) follows from the 200 operating days, the probability of the temperature and the weight of the pax level 51
6.8 Cancelled trips for the benchmark schedule of the Lelystad depot under the scenario with average pax and a temperature of 0°C 54
6.9 Cancelled trips for the benchmark schedule of the Lelystad depot under the scenario with average pax and a temperature of -5°C 55
6.10 Performance of the benchmark schedule per scenario day (mean over 1,000 Monte Carlo runs), for Apeldoorn (red rows) and Lelystad (green rows). The scenario is described by the temperature and the pax level 56
6.11 Annual performance of the benchmark schedules, weighted over all scenarios by the expected number of days per year 57
6.12 Completion rate and net change in cost compared with the benchmark schedule, without driver cost. All values are percentages: the changes are percentages of the annual operator cost of the benchmark, and the planned schedule cost is compared with the reference described above 60
6.13 Reliability and delay performance for depots Apeldoorn (red rows) and Lelystad (green rows) across schedules 61
6.14 Efficiency performance for depots Apeldoorn (red rows) and Lelystad (green rows) across schedules 63

6.15 Operator cost performance for depots Apeldoorn (red rows) and Lelystad (green rows) across schedules, without driver cost. The planned cost is compared with the average of the -10 and $+10$ schedules, the cancellation cost with the benchmark schedule . . . 63
6.16 Net change in cost compared with the benchmark schedule, including driver cost. All values are percentages of the annual operator cost of the benchmark without driver cost; the planned schedule cost is compared with the reference described in section 6.4.1. The cancellation cost and the passenger delay cost do not change by including driver cost, so the latter is only given in table 6.12 (the overall change includes it). The range is that of the overall change for fines of €0 to €40 per cancelled trip and idle hours valued at 0% to 50% of the rate for driving hours 65

7.1 General findings derived from the case study 71

C.1 Planned schedule cost and overall net change in cost when the planned schedule cost of the benchmark is used as reference (without correction), as a percentage of the annual operator cost of the benchmark without driver cost 83

Nomenclature

Abbreviations

Abbreviation	Definition
CV	Cross-validation
DES	Discrete event simulation
EBS	EBS Public Transportation
EC	Energy consumption
HVAC	Heating, ventilation and air conditioning
KNMI	Royal Netherlands Meteorological Institute
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MNR	Multivariate non-linear regression
OLS	Ordinary least squares
pax	Passengers
PDP	Partial dependence plot
P5 / P50 / P95	5^{th} percentile / median / 95^{th} percentile
RMSE	Root mean square error
SoC	State of charge
SoH	State of health
VoT	Value of time

Symbols

Symbol	Definition	Unit
c^{fine}	Fine per cancelled trip	[€]
c^{fixed}	Fixed cost per bus per day	[€/day]
c^{variable}	Variable cost per kilometre	[€/km]
c_i^{saved}	Variable cost saved by cancelling trip i	[€]
$C_{p,s}^{\text{passenger}}$	Passenger delay cost of schedule p in scenario s	[€]
$C_{p,s}^{\text{disturbance}}$	Operator cost of cancelled trips of schedule p in scenario s	[€]
C_p^{schedule}	Planned schedule cost of schedule p	[€/year]
ΔC_p^{pass}	Net change in passenger cost of schedule p	[€/year]
ΔC_p^{op}	Net change in operator cost of schedule p	[€/year]
ΔC_p^{all}	Overall net change in cost of schedule p	[€/year]
d_p^{service}	Service distance of schedule p	[km/day]
d_p^{total}	Total driven distance of schedule p	[km/day]
D^{op}	Number of operating days in a year	[days]
D_s	Expected number of days per year of scenario s	[days]
$E_p^{t,\text{drive}}$	Service time as a share of driving time	[–]
$E_p^{t,\text{total}}$	Service time as a share of total vehicle task time	[–]
E_p^d	Service distance as a share of total driven distance	[–]
h_i	Headway between trip i and the next viable trip	[min]
m_i	Missed income of cancelled trip i	[€]
$N_{p,s}^{\text{canc}}$	Number of cancelled trips of schedule p in scenario s	[–]
p	Schedule	[–]
p_0	Benchmark schedule	[–]
$PW_{p,s}$	Passenger delay of schedule p in scenario s	[h]
q_i	Number of passengers on trip i	[–]
$R_{p,s}$	Trip completion rate of schedule p in scenario s	[%]
s	Scenario	[–]
t_p^{service}	Service time of schedule p	[h/day]
t_p^{driving}	Driving time (service and dead-heading) of schedule p	[h/day]
t_p^{total}	Total vehicle task time of schedule p	[h/day]
v_p	Number of buses (vehicle tasks) required for schedule p	[–]
VoT^{IV}	In-vehicle value of time	[€/h]
w_s	Weight of scenario s	[–]
x_{ij}	Standardised value of factor j for trip i	[–]
y_i	Energy consumption per kilometre of trip i	[kWh/km]
β_0	Intercept of the regression model	[kWh/km]
β_j	Coefficient of the linear term of factor j	[kWh/km]
β_{jj}	Coefficient of the squared term of factor j	[kWh/km]
β_{jk}	Coefficient of the interaction between factors j and k	[kWh/km]
ε_i	Unexplained part of the energy consumption of trip i	[kWh/km]
λ^{wait}	Multiplier for out-of-vehicle waiting time	[–]
λ^{arr}	Multiplier for unexpected late arrival	[–]
π_s^{temp}	Probability of the temperature of scenario s	[–]
π_s^{pax}	Weight of the pax level of scenario s	[–]

1

Introduction

Research has clearly shown the negative impact of climate change due to greenhouse gas (GHG) emissions (IPCC, 2021). Transportation delivered almost a quarter of European GHG emissions in 2022 and is therefore subject to efforts to decrease those emissions (European Parlement, 2024). European legislation has set a target for the reduction in emissions from heavy duty vehicles, including public transport buses. The regulatory framework requires GHG emissions of battery electric buses used in public transport to decrease by 45% in 2030 and up to 90% in 2040 (European Commission, 2024). Within bus operations, the most important aspect of the emissions come from the fuel consumed, due to the relative long life time of the equipment itself (Szczurowski et al., 2022; W. Zhang et al., 2024). Improving sustainability for bus operators thus means, moving from fossil fuels to low or zero emission energy sources. The Dutch public transit sector has already agreed that from 2025 onward newly bought buses, and in 2030 all operated buses should be zero-emission (BAZEB, n.d.).

The use of battery electric propulsion systems in public transport has seen a remarkable rise over the past decade. The amount of electric buses operating in the Netherlands went from 52 in 2015 to 2766 in 2025 (CROW-KpVV, 2026). However, the implementation of electric buses also created new challenges. Problems arise due to limited range, high initial costs and the use of new charging methods and infrastructure (Alamatsaz et al., 2022; Chau et al., 2024; Wang et al., 2025). The limited range and placement of charging infrastructure (if grid connections are available in the first place) have a high impact on strategic planning. The placement of infrastructure is adjusted on the expected range of the buses and needs to be completed before operations start. Once installed, planning and operations are bound to the existing infrastructure. Next to the limited driving range itself, the fluctuations in energy consumption (EC) further decrease available range. This uncertainty requires energy redundancy in the planning phase to prevent buses coming to a standstill during operations (Alamatsaz et al., 2022; Perumal et al., 2022).

1.1. Scope

The planning process of buses in public transport features several planning steps on the strategic and tactical level. The adaptation of battery electric buses meant the inclusion of additional steps for the design of charging infrastructure and charging scheduling. The process is shown in figure 1.1. The diagram is derived from Perumal et al. (2022). The dark green steps indicate the scope of this research.

Choices that impact the fleet's ability to fulfil the complete set of routes are made during the fleet scheduling step. With the introduction of electric buses, an additional step is required to plan charging events during the day. These two are often combined within optimization literature and are called the electric vehicle scheduling and charging problem (EVSCP). Although this study does not explicitly model the optimization, it is still the main focus in the context of the planning steps. Although investment decisions for the bus fleet are made in advance, the option to invest in new buses within the optimization is kept, to show the influence of adjustments on the number of buses that would be required. The same holds true for the recovery step. The simulation only includes simple recovery actions for buses that

would otherwise run out of battery, and not advanced decision making at that level. However, because the results of choices in earlier planning phases are analysed on the operational level, it is still included as part of the scope. The estimation of EC is also part of the scope, as it provides the input to calculate the energy that a bus uses in the simulation.

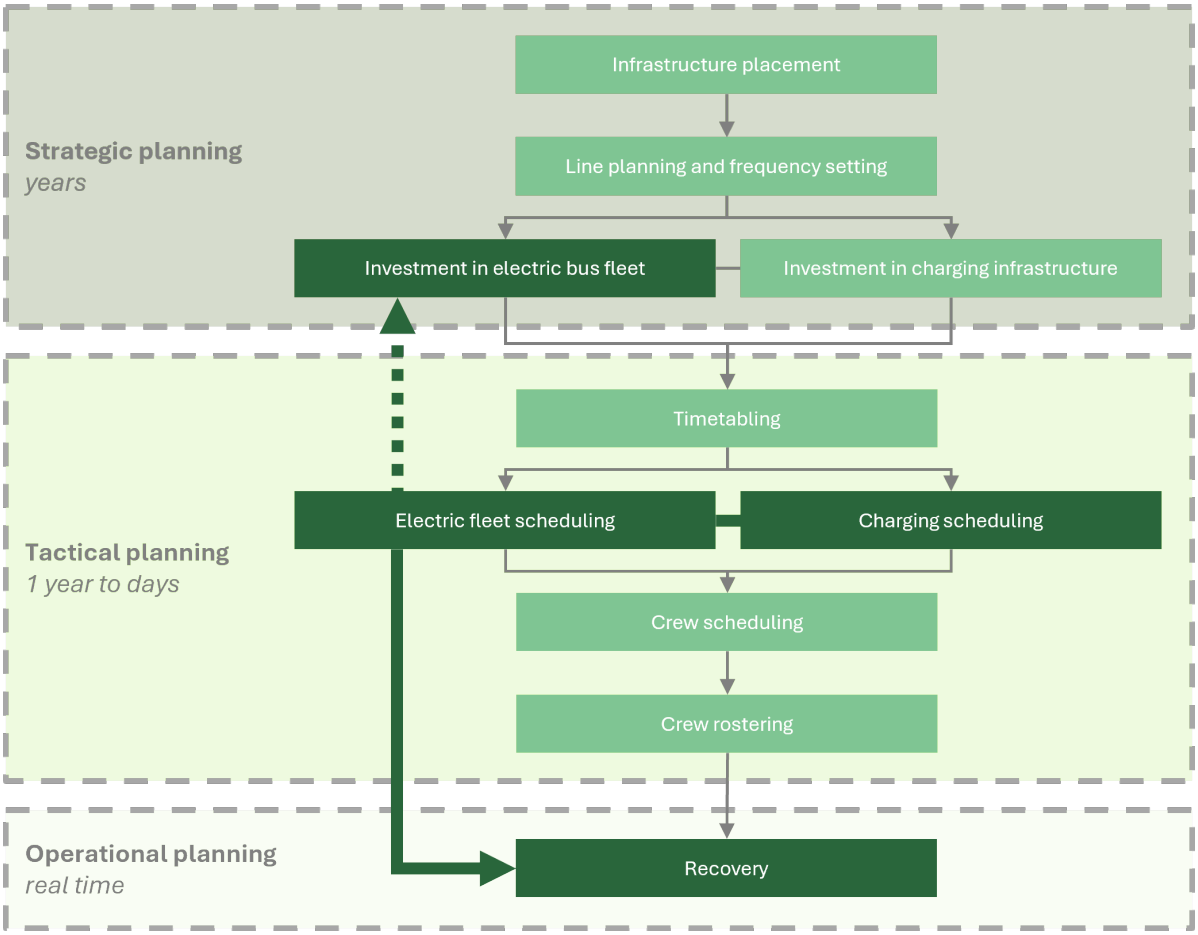


Figure 1.1: Electric bus planning process, derived from Perumal et al. (2022), with the steps that fall into the scope and the relationships of interest indicated by a dark green colour

EBS Public Transportation (from here on referred to as EBS) is a Dutch transport operator with over 500 battery electric buses (from here on referred to as simply buses) throughout the Netherlands. In their transition to zero-emission transportation, they encountered the challenges mentioned above. These challenges have an impact on service provided by the company. Desaulniers and Hickman (2007) and Ibarra-Rojas et al. (2015) state that the main focus of a transport company is to provide high quality service while minimizing costs. Service quality is determined by different aspects that influence rider satisfaction (Van Oort, 2011). For an order in which aspects are the most important to fulfil, studies often refer to the *pyramid of transit needs* by Peek and Hagen, 2002, an adaptation of the original Maslow's pyramid of needs. The inclusion of reliability as a fundamental layer indicates that it must be adequately fulfilled, before thinking about improving other aspects of public transport.

For this research, reliability is narrowed down to solely service reliability. This is, on its turn, specified as the ability of a bus to complete a scheduled trip on a given day with enough energy remaining to reach a charging point - where the EC of previous trips by that bus on that day should also be taken into account. This is referred to as the trip completion rate. On its own, however, it says little about the circumstances in which the trips occurred. Robustness extends trip completion to this dimension. According to Lusby et al. (2018) robustness refers to the extent to which a predetermined plan can deal with unforeseen variations without requiring structural changes. A schedule that leads to a reliable service (i.e. all trips completed) under a wide range of conditions can be seen as a robust schedule.

The scope of this research follows from the above. The study looks at the tactical level, in which the vehicle schedule (the assignment of trips to buses, including the charging events) is made for a fixed timetable and fixed charging infrastructure. The planning value for the EC is the lever that is studied: it determines how much energy redundancy the vehicle schedule contains. The consequences of this choice are assessed on the operational level, by simulating how the schedule performs on days with different temperatures and passenger numbers. Several related topics are deliberately left out. The level of service, in the form of frequencies and timetables, is not changed. The placement and dimensioning of charging infrastructure, grid congestion and crew scheduling are also outside the scope, although the potential impact of fleet scheduling on the latter is shortly assessed. The effects on driver schedules are illustrated with an estimate that uses schedule efficiency. The case study consists of two depots in the Netherlands, each with its own schedule, so that the interaction between depots is not studied either.

1.2. Research objective and contributions

The relevance of this study is both scientific and societal. Scientifically, EC uncertainty is acknowledged in the literature on electric bus scheduling, but the effect of the EC value that is assumed during planning on the robustness and efficiency of the resulting schedules has hardly been quantified (see chapter 2). For EBS and other operators, the question is directly relevant. A planning value that is too optimistic leads to cancelled trips on cold days, with consequences for passengers and for the subsidy that is received. A planning value that is too conservative leads to extra buses and extra kilometres and affects the operator directly in terms of cost associated with the schedule. In the long term, that may also lead to higher public spending, as part of the revenue comes from subsidies. Operators currently have to find the balance based on experience and rules of thumb, as the numbers to support the decision are lacking. With the transition to zero-emission buses largely expected to be completed in the coming years, the EC value that is assumed in planning is a recurring decision for every new concession and every new timetable.

The paragraphs above inform us about the importance of service reliability for a passenger, and therefore schedule robustness for a transport operator. It also mentions the goal of minimizing costs and in order to do so, a company needs to use their resources efficiently. That is where robust schedule starts to compete with an efficient schedule. If, for example, a schedule is designed to handle extreme conditions where EC is 100% higher, it means that buses would have abundant energy in normal conditions. While this trade-off is quite intuitive, little is known about the quantitative effects of this trade-off. This research aims to facilitate the discussion on the desired level of robustness by quantifying the effects of a change in robustness on the costs of operations. That resulted in the following main research question: *How does the energy consumption value that is assumed in the scheduling of electric buses shape the trade-off between robustness and efficiency, for public transport facing energy consumption uncertainty?* The EC model is the link between the planning and the operations in this study: it is used in the simulation to calculate the energy that a bus uses on every trip, given the conditions of a scenario, so that the consequences of a planning value can be assessed.

The approach designed to fill this research gap has three main contributions to the scientific field:

1. the development and validation of a data-driven EC model, based on a full year of real trips of the Dutch region IJssel-Vecht, that quantifies the influence of external factors and is validated both on trip and on vehicle task level, which is the level at which it is used for scheduling;
2. a scenario-based simulation framework that combines the EC model, scenarios for temperature and passenger numbers and the unexplained variation of the EC model (Monte Carlo) to assess the robustness of electric bus schedules under EC uncertainty;
3. a quantification of the trade-off between robustness and efficiency, by comparing schedules that are generated with different EC planning values under the same scenarios and expressing the outcomes in trip completion, societal cost and operator cost.

For EBS, and for public transport operators in general, the study contributes to practice in a number of ways. It gives an indication of what energy redundancy costs and what it brings, for passengers as well as for the operator. It provides a framework that can be reused with the data that an operator already collects, to support decisions on planning values for new concessions or timetables. Finally, it shows

how charging infrastructure, in particular fast charging at a hub versus slower charging at the depot, changes the consequences of a lack on robustness during extreme events.

1.3. Outline

The remainder of this thesis starts with a literature review of the state of the art in chapter 2. This is followed by an explanation of the general methodology, including the sub-questions, in chapter 3. The methodology is applied to a case study in the IJssel-Vecht region, which is introduced in chapter 4. Chapter 5 describes and analyses the data that is used for the case study. It can be seen as an intermediary step between the methodology and the results, and features aspects from both. Logically, the chapter that follows, chapter 6, contains the results. The outcomes are compared with other studies, and their limitations are discussed, within the sections of that chapter. Last, the research questions are answered in chapter 7, which also formulates the scientific contributions, the general findings and the recommendations for practice and for further research.

2

Literature review on state-of-the-art

The literature review aims to provide state-of-the-art on three topics, being energy consumption estimation, robustness and fleet scheduling in public transport. The main sources of information are Google Scholar, WorldCat Online via the TU Delft library and snowballing in articles provided by the previous sources. For the WorldCat searches, the queries and results are presented in table 2.1. For Google Scholar additional terms - gathered from the articles provided by the WorldCat search - were used alongside the words shown in the table. The results were too plentiful to derive a meaningful statistic. In general only the first 30 results sorted on relevance were scanned and read if applicable. The literature on each topic will be discussed separately in the sections below. Peer-reviewed and recent articles and searches on key words were preferred. However, when a search gave too many results or too few, the filters were adjusted, so that the search could be reproduced, but also remained doable within the timeframe of this study. That is why the search on robustness assessment is based on the title instead of key words, and why the period that was searched differs per topic. The filters that were applied are listed in table 2.1.

2.1. Energy consumption estimation

2.1.1. Modelling techniques

Estimating energy consumption for battery electric vehicles and more specifically buses has gotten increased attention from researchers. Two recent reviews by X. Zhang et al. (2024) and Jahic et al. (2023) paint a good picture of current modelling techniques and developments therein. Each technique has its own strengths and weaknesses:

- **Empirical** models are the simplest and fastest way to get a rough estimation of energy consumption, but should not be used for advanced decision making. Generally, an average number for consumption is multiplied by the distance that has to be covered.
- **Physical** models rely on vehicle dynamics the mechanics of additional subsystems, such as the

Table 2.1: Literature search results

Subject	Aspect	Query	Filters	Results	Relevant
EC modelling	Keywords	*energy consumption* AND (*electric bus* OR *electric bus*) AND (*public transit* OR *public transport*) AND *estimation*	peer reviewed 2020–2026	37	8
Robustness assessment	Title	*robustness public transport*	peer reviewed	24	15
Electric vehicle scheduling and charging	Keywords	(*"charging problem"* OR *"scheduling problem"*) AND *electric bus* AND *public transport*	2015–2026	15	9

heating, ventilation and air conditioning (HVAC) for the power consumption calculations. The amount of parameters to include and the available data often require simplifications. For example, calculating drag from wind is almost impossible as wind speed and direction are constantly changing. Still, the calculations with those simplifications in mind can lead to accurate estimations.

- **Regression or statistical** models, together with the remaining two types of models, are referred to as data-driven models. Here one or multiple explanatory variables are used to determine the function that explains the energy consumption. A further distinction can be made between linear, non-linear and non-parametric regression.
- **Machine learning** models: although the previous category of models uses machine learning as well, both reviews set the regression models apart from other machine learning models. Machine learning models can automatically adjust parameters, learning from the data and therefore requires no previous knowledge or initial parameters. This can lead to better estimations, because human bias is prevented. However, the model is not transparent and therefore often referred to as a black-box.

X. Zhang et al. (2024) dive further into neural networks as a distinct category within Machine learning models. They also include the category of Hybrid models (grey-box), where machine learning methods are combined with physical (white-box) models. The accuracy is in most cases expressed in mean average percentage error (MAPE). The lower the MAPE of a model, the higher the accuracy. In order to use the outcomes from the model as input for scheduling, the accuracy should be as high as possible. This is because higher accuracy requires lower margins to cover uncertainty.

Next to the type of model, the granularity also plays a large role in determining the right approach. Two main options are macroscopic and microscopic. Macroscopic refers to models that focus on complete networks or fleets and are often used for strategic or tactical planning. Microscopic calculate energy consumption per unit of time, often as detailed as 1Hz (once per second). This can be used for planning purposes as well, although the amount of required data and computations would be greater. A more suited application is real-time fleet management due to live estimation.

2.1.2. Influential parameters

Besides the variety in models used, the literature also shows great diversity in the factors that are included in the models. A complete overview of the parameters used by studies in this section can be found in table 2.2. (Belloni et al., 2025) tracked the consumption of each subsystem of electric bus and gathered experimental data on different scenarios. The electric motor and the HVAC (heating, ventilation and air conditioning) used the most energy – about 45% and 34% respectively. Power steering accounted for 11%, which may indicate the importance of corners during the route. Their sensitivity analyses show how a well-trained driver can consume up to 50% less energy. They also divided the EC into different electrical systems of the bus. It should be noted that the route characteristics are not described in detail and the temperatures were very high during their 18 day sampling period, averaging 34°C. However, it does give insight into the systems that require energy during operations.

Ji et al. (2022) use a multivariate non-linear regression model and showed *distance* has a large influence, as well as *curb weight*. The factor temperature showed a non-linear relationship with energy consumption. There seems to be an optimum at around 24°C and the EC rises as the ambient temperature moves further away from that optimum. In their data they had a broad range of temperatures available, ranging from -27 to +35°C. The lowest records resulted in an extra TEC of almost 50%. The MAPE of the model is about 12%. Abdelaty and Mohamed (2021) used a two step approach where they first simulated the energy consumption through a vehicle dynamics model and assumed auxiliary power demand. The simulation provided fictional EC rates, but were verified with an accuracy of about 5% against real world data. Second, they applied multivariate linear regression (MLR) where they included nine independent parameters. The *road grade* had the highest impact, followed by *initial SoC*, *road condition* and *HVAC*.

Li et al. (2021) used a random forest (RF) technique and although it achieved a low accuracy with an MAPE of about 15%, it used *number of traffic lights* as a factor, which was not included in the other articles. After running a variable importance analysis on their RF model, the *trip length*, followed by

number of stops and *number of traffic lights* came out as most important factors. Chen et al. (2021) combine data-driven models in the form of a long short-term memory (LSTM) and artificial neural network (ANN) with insights from a physical model for their grey-box model approach. Their goal is to improve prediction accuracy by incorporating vehicle specific power, derived from their physical model, as an additional factor in their LSTM model. The model that included *speed*, *acceleration*, *vehicle specific power*, *weather* and *route characteristics* achieved the highest accuracy with a MAPE of 2.8% per second, which is 4.2% when compared on trip level. Z. Zhang et al. (2024) show how one could improve the interpretation of pure black-box models by doing a variable importance analysis and a Shapley analysis. They analysed 25 factors that could influence energy consumption and reviewed their importance within three models, being XGBoost, RF and LightGBM. Although the importance differs per model, there are four aspects in every top 10. *Temperature outside*, *average speed*, *congestion index* and *average acceleration pedal*. The latter is one of the indicators for driving behaviour, while the others fall within environmental or traffic conditions. Their Shape value analysis states HVAC and travel time as the most important factors. Their LightGBM model using 25 factors managed to bring the MAPE as low as 4.4%.

Table 2.2: Overview of techniques and parameters used in the literature

	Chen et al. 2021	Ji et al. 2022	Zhang et al. 2024	Abdelaty and Mohamed 2021	Li et al. 2021	Belloni et al. 2025	This study
Granularity	1 Hz	trip	trip	n.a.	trip	trip	trip
Technique	LSTM and ANN	MNR	Various data-driven	Simulation and MLR	RF	Sensitivity analysis	MNR
Vehicles	1 type	3 types	2 types	1 type	n.a.	1 type	2 types
Routes	12 routes	3 routes	3 routes	n.a.	n.a.	4 routes	>40
Vehicle parameters							
Vehicle mass	x	x		x		x	
SOC at departure			x	x			x
HVAC		x	x	x			
Operational parameters							
Driving direction			x		x		
Speed	x		x	x			
Average vehicle speed		x	x		x	x	x
Acceleration	x		x	x			
Driving style			x	x		x	
Time of day			x		x		
Day of week			x		x		
Passenger load				x			x
Route parameters							
Distance		x	x	x	x		x
Gradient	x			x		x	
Number of stops			x	x	x		
Number of intersections					x		
Road type	x						
Environmental parameters							
Traffic			x	x		x	
Temperature	x	x	x	x	x	x	x
Weather conditions	x			x	x		x

Not all variables that are used in the literature are included in this study, mainly because the time to obtain all factors was limited and because the factors should be available to operators in general. The road gradient, the number of stops and intersections and traffic are therefore not part of the model. The trip distance is not used as an explanatory variable either. When the timetable is fixed, the distance of a trip does not change, so the preference is to look at the EC per unit of distance.

In short, temperature is named as an important factor by most reviewed studies, but the studies do not agree on the other factors, so it is not clear beforehand which factors matter most. Regression models are preferred in this study because they are transparent and fast to compute. That is relevant, as the model has to be used for every trip in many simulations, and the effect of each factor has to be interpretable to select scenarios. The model in this study differs from earlier work in the combination of factors that it includes. In addition, the reviewed studies report the accuracy of their models per trip or at an even finer resolution, whereas fleet scheduling asks for accuracy over the combined trips of a bus. This builds up towards the first scientific contribution of this study.

2.2. Robustness in public transit

The introduction of this work already mentioned the general definition of robustness provided by Lusby et al. (2018). In short, the ability of a system to cope with uncertainty. A deviation from expected procedures is referred to as a disturbance. A disturbance can be, for example, lower (than expected)

temperatures, a collision or power outages. Both Lusby et al. (2018) and Ge et al. (2022) add that, while there is a general understanding of robustness, the detailed operationalisation of robustness differs strongly between applications. Getting a solid understanding of how robustness is measured and assessed within public transport is therefore the main goal of this section in the literature review.

The relevant work on robustness can be roughly divided into two categories: the analysis on network level, with a focus on link and node failures during disruption, and optimization under uncertainty. The latter serves as a method to improve robustness in the strategic and tactical planning phase by incorporating uncertainty and redundancy in the optimization.

2.2.1. Network robustness

Multiple studies use graph theory to measure network performance and impact of node and link failures. The goal is to identify potential bottlenecks during disruptions or to indicate critical links in the network. This is for example done by Aparício et al. (2022) and Cicchini et al. (2024). For a complete overview of public transit network indicators derived from graph theory, see the recent survey from Palk et al. (2026). Other studies base network performance on passenger indicators. Yap (2014) includes a societal cost benefit analysis in his work. Using the multimodal rail network of the The Hague-Rotterdam metropolitan area, he first identified disturbances and the possibility of their occurrence. OmniTRANS is chosen as the model to provide passenger assignment. Having both the probability of disturbances and the amount of passengers on link level, a selection is made for most vulnerable links. With the value of time (VoT) and perceived (dis)utility, the effects for passengers during disruptions are quantified. Ultimately the infrastructure investments that would create a more robust network and their associated costs, are set of against the welfare benefits (e.g. shorter travel times and uncompleted trips). Cats and Jenelius (2015) lay more emphasis on spill-back effects of disruptions. Using the BusMezzo tool, they simulate having increased capacity on alternative links to deal with unexpected disruptions. In another study from the same authors, they analyse the impact of adding or expanding routes in the network for robustness (Jenelius & Cats, 2015). Two further studies are relevant for the way in which robustness is quantified. Cats (2016) assesses the robustness value of public transport development plans in Stockholm through a full scan of link failures, and reports both the average and the worst-case deterioration of performance. He shows that neglecting abnormal operations in the appraisal of an investment leads to an underestimation of its benefits. Cats et al. (2016) add that the probability of a disruption (exposure) should be part of the assessment. For a multimodal network in the Netherlands, they define the criticality of a link as the probability of a failure multiplied by the resulting welfare loss, and show that ignoring probabilities leads to a different prioritisation of links. Both studies support an approach in which the consequences of disruptions are weighted by their probability, and in which the worst case is reported separately.

2.2.2. Optimizing under uncertainty

Apart from infrastructure adjustments or line planning, a public transport operator has more tools during the planning phase to provide robust service. In timetabling for example, the operator can add buffer times between trips to prevent delays from one trip propagating onto the next. This can prevent cost associated with delays. Ge et al. (2022) point out how uncertainty is dealt with in public transport optimization. Stochastic programming assumes that uncertain parameters follow a known distribution. This often translates to assigning probabilities to scenarios. Allowing a great number of scenarios comes at the cost of computing efficiency. Robust optimization on the other hand requires only ranges or deviations from nominal values. However, they tend to lead to conservative outcomes. An upcoming method that aims at combining elements of the two, is distributionally robust optimization (DRO). It uses an ambiguity set that captures multiple possible distributions, where stochastic programming assumes just one probability distribution to be true. Momen et al. (2025) apply this principle for charging infrastructure design under sparse energy-consumption data. Their ambiguity set is constructed around the empirical distribution, while a displacement budget controls how far alternative distributions may deviate from the observed data and chance constraints specify the required probability of feasibility. Their results quantify the cost of robustness for the Rotterdam bus network: ignoring the variation in EC results in infeasible trips in up to 55% of the scenarios, while designing the infrastructure for the worst-case EC increases the cost by 67% compared to a design based on average estimates. The data-driven DRO model reduces the cost by 28% compared to the model with box uncertainty, while

maintaining reliability.

On the topic of bus scheduling, Bie et al. (2021) use data from real trips to derive the stochastic volatility for travel time and EC. EC is estimated through a regression model that considers SoC at the start of a trip, travel time and average temperature during the trip and they use the unexplained variance as probability within their optimization. Travel time is incorporated as a discrete random variable with steps of one minute. They optimize three functions: 1) departure delays, 2) total EC, and 3) costs. De Bruin et al. (2023) have a similar approach to obtain travel time distributions, but incorporate them in a discrete-event simulation (DES). They define normal and busy scenarios and preserve dependencies between trips within the same day. They include the robustness assessment within their algorithm rather than using it as a post-optimization assessment. Final deterministic and stochastic schedules are subsequently compared using Monte Carlo simulation, with punctuality, mean and percentile lateness, fleet size and operating cost as evaluation criteria. Tang et al. (2019) propose robust scheduling strategies for electric buses under stochastic traffic conditions. Their static model includes a buffer-distance strategy: a safety margin on the range of the buses that prevents en-route breakdowns when trips take longer than planned. The results for bus lines in Beijing show that breakdowns can be avoided while maintaining cost efficiency. The buffer is conceptually close to energy redundancy in the EC value, but is not varied and quantified as a trade-off. Manzolli et al. (2025) take the EC uncertainty into the charging strategy of an electric bus fleet. Their robust optimisation model protects against energy consumption deviations, while also accounting for battery ageing and time-of-use tariffs. In their case study, the cost reduction compared to a business-as-usual charging strategy is 12% for the robust model, against 37% for the deterministic model.

Yap and Van Oort (2018) explore the potential gains and losses from robust driver scheduling amid disruption for a case study on The Hague's light rail network. They develop a framework in which the passenger and operator cost of a large disruption are explicitly expressed as a function of the complexity of the driver schedule. A more complex schedule, in which a driver works on several vehicles or lines during a shift, reduces operator cost under normal operations, but results in longer recovery times and less effective rescheduling once a disruption occurs. The passenger cost are based on passive data and valued with the VoT and a weight for the perceived waiting time. The operator cost include personnel overtime and fines for early and late departures and cancelled trips. They find that, for the case study, the total benefit of a less complex schedule (the sum of the passenger and the operator benefit) is larger than the additional schedule cost, whereas the benefit for the operator alone is smaller. The study is the closest example of how a trade-off between efficiency and robustness can be quantified, but it addresses driver schedules and a single, large, non-recurrent disruption. The indicators for robustness used in the literature are summarized in table 2.3.

Table 2.3: Literature overview of robustness perspectives, subjects, and indicators

Perspective	Subject	Indicators	Articles
Network	Connectivity and accessibility	Largest connected component; network efficiency; path length; degree/betweenness; link criticality; degradation rapidity	Aparício et al. (2022), Cats & Jenelius (2015), Cicchini et al. (2024), and Jenelius & Cats (2015)
Passenger	Ability to complete itineraries as expected	Affected, delayed and cut-off passengers; additional waiting and in-vehicle time; transfers; denied boarding; welfare loss	Cats & Jenelius (2015), Jenelius & Cats (2015), Yap (2014), and Yap & Van Oort (2018)
Operator	Ability to operate all scheduled trips with available vehicles, time and energy	Trip completion; delay propagation; slack; vehicle utilization; reserve capacity; energy-feasible trips; cost and reliability	Bie et al. (2021), De Bruin et al. (2023), and Yap & Van Oort (2018)

To conclude, the reviewed studies assess robustness from different perspectives, but for this study the effects on passengers are most relevant. They are expressed as the societal cost of additional waiting time, valued with the value of time, and complemented by the trip completion and the cost for the operator. That the consequences of disruptions should be weighted by their probability, and the

worst case reported separately, supports a scenario approach. The trade-off between robustness and efficiency has only been quantified in a few studies. Yap and Van Oort (2018) is the closest example, but concerns driver schedules and a single disruption. It is used as inspiration, and this study can be seen as an extension of it to vehicle scheduling with multiple scenarios that are weighted with their probability. The other studies quantify the cost of robustness for charging infrastructure design (Momen et al., 2025), charging strategies (Manzolli et al., 2025) or a buffer distance (Tang et al., 2019), but not for the EC value that is used in planning. To understand which variables and parameters are important for fleet scheduling, the next section turns to the planning problem itself.

2.3. Fleet scheduling and charging optimization

The allocation of electric buses to trips is referred to as the electric vehicle scheduling problem. The main constraint is that every trip needs to be assigned to exactly one bus. The goal is to minimize cost, although other objectives can also be incorporated (see Behnia et al. (2024) for an elaborate overview). With the introduction of electric buses, charging scheduling has also found its way into public transport planning. Although there are articles that focus on charging scheduling as an independent planning step (e.g. Manzolli et al. (2025)), the focus for this state-of-the-art review is on the integrated electric vehicle scheduling and charging problem (EVSCP). That also means that studies that integrate other planning steps, such as timetabling (Gao et al., 2025) or charging infrastructure placement (Yıldırım & Yıldız, 2021), fall out of the scope of this review.

While a search has been performed as stated in table 2.1, together with additional searches in Google Scholar, it should be noted that the main source of information is a mere total of five literature reviews - all written within the last five years (Behnia et al., 2024; Lim et al., 2023; Perumal et al., 2022; Wang et al., 2025; L. Zhang et al., 2023). Perumal et al. (2022) take a broad perspective and analyse the introduction of electric buses on all planning steps. L. Zhang et al. (2023) focus on the solution methods and go into detail about integration with timetabling and crew scheduling. Lim et al. (2023) dive into the constraints associated with charging infrastructure. This can for example be a fixed amount of chargers on a location or limitations from power grid operators. Their work shows the extra complexity associated with these constraints. Behnia et al. (2024) have perhaps the most elaborate survey, when it comes to highlighting all the different considerations (e.g. objective, constraint source) for electric bus scheduling.

The evolution from traditional vehicle scheduling problems (VSP) into EVSPs meant the addition of new constraints. The addition of more complexity to an already complex problem pushed the search for heuristics that can decrease computational time. Important sub-categories in the ESVP are single or multiple depot and single or multiple vehicle types, with multiple types again leading to more computational demand. The following paragraphs will discuss the approaches taken by recent studies. A complete overview of solution approaches and operational characteristics for the selected articles is shown in table 2.4 and 2.5.

Table 2.4: Comparison of problem characteristics and solution approaches

Source	Problem	Objective	Mathematical formulation	Solution method	Metaheuristic	Network representation	Trips
Wen et al. (2016)	EVSCP	Min. fleet and distance	SPP	Metaheuristic + post-optimization	ALNS + SA	Event-based graph	500
Yao et al. (2020)	EVSP	Min. annual cost	MCFP	GA + bin-packing heuristic	GA	Time-space graph	931
Zhang (Le) et al. (2021)	EVSCP	Min. cost and degradation	SPP	Exact (branch-and-price)	-	Directed connection graph	218
Wu et al. (2022)	EVSCP	Min. cost and peak load	MCFP	Exact (branch-and-price)	-	Time-expanded graph	379
Zhang (Aijia) et al. (2022)	EVSP	Min. total cost	MCFP	Stand-alone heuristic	ALNS	Directed-connection graph	620

Starting with exact methods, both L. Zhang et al. (2021) and Wu et al. (2022) use branch-and-price to find a solution to the EVSCP. The difference is that the former use a formulation built on the set partitioning problem and the latter is based upon a multi-commodity flow problem. L. Zhang et al.

Table 2.5: Comparison of operational characteristics

Source	Multiple depot	Multiple vehicle types	Multiple charger types	Partial charging	Charging curve	Charging rate	Limited chargers
Wen et al. (2016)	multiple	single	single	partial	linear	0.8 min/km	no
Yao et al. (2020)	multiple	multiple	single	full	linear	240 kW	no
Zhang (Le) et al. (2021)	single	single	single	full	non-linear	variable	yes
Wu et al. (2022)	multiple	single	single	full	linear	200 kW	partly
Zhang (Aijia) et al. (2022)	multiple	multiple	single	partial	linear	72–130 kW	no

(2021) implement a non-linear charging curve to better capture real-world charging practices. Wu et al. (2022) focus on the application of power grid constraints and time-of-use pricing for energy. The power grid constraints prevent the buses from taking too much power at the same time, while the inclusion of time-of-use energy pricing discourages charging during peak hours.

Then over to the heuristic approaches. Yao et al. (2020) use a genetic algorithm (GA) with a bin-packing approach, where they 'load' sets of sorted trips into vehicle tasks. The GA finds the optimal sorting of trips. They include multiple electric bus types, where each type differs in battery capacity and EC. Wen et al. (2016) introduce partial charging for the EVSCP. They propose an adaptive large neighbourhood search (ALNS) with post-optimization in the form of a set partitioning problem (SPP). The ALNS works by destroy and repair operations, where a part of the column (vehicle task) is replaced by other events (trips). With a form of simulated annealing (SA), the replacement of trips is done according to a temperature indicator. After a fixed amount of iterations the SPP picks out the best combination of columns. They compared the model for small instances against a mixed integer programming (MIP) formulation and achieved similar results in a fraction of the computing time. Their directed acyclic graph can best be categorised as an event-based graph, where they use exact start and end times for trips and charging. The difference between an ordinary event-based graph is that they add a copy of a charging node before and after a trip. With that, they prevent having cycles in the graph, but also adds the assumption of limitless charger availability. A. Zhang et al. (2022) too, accommodates partial charging policies and use an ALNS to solve the problem, but without post-optimization. They extend the model with multiple vehicle types. Within the repair part of their ALNS, they add to option to upgrade the bus' battery size, if larger models are available. A MIP following a multi-commodity flow problem (MCFP) formulation is presented and used to test the optimality gap up to instances with 80 trips. After that, the CPLEX solver cannot obtain the optimal solution within 6 hours, which makes a clear argument for using heuristics.

A study that fell out of the scope, but still could be used as inspiration, is the work from Fan et al. (2024). The researchers propose a non-linear EC model, to incorporate route characteristics, such as slope or passenger load, into the optimization. It is roughly based on longitudinal vehicle dynamics and allows for heterogeneity in EC explored earlier in this chapter.

To conclude, the electric vehicle scheduling and charging problem is tackling increasingly larger problems, but also becomes increasingly complex in its setup. Developing an own optimisation model was not possible within the time limit of this research. The literature is nevertheless useful to understand which variables and parameters are important for fleet scheduling, such as the EC rate, the number and power of chargers and the charging method.

2.4. Conclusions from the reviewed literature

This review touched upon the state-of-the-art of several topics related to electric bus scheduling. The literature on EC modelling shows that it is possible to estimate EC with a data-driven approach. The wide variety of models and outcomes, however, shows that identifying the most influential factors is not straightforward. This especially holds for black-box models. Regression models, on the other hand, are the most transparent method, but require assumptions on the relationship between EC and the factors involved. In public transport, robustness can take various shapes and sizes. The review shows that there is increasing attention to the effects experienced by passengers, underscoring that

part of the trade-off that a transport operator has to consider. Stochastic and robust optimization come forward as methods to account for travel time or EC uncertainty, but those methods require knowledge on scenario probability or parameter distributions. Normal optimization approaches for the EVS(C)P, meanwhile, are shifting from exact methods towards metaheuristics. Due to the increasing complexity, exact methods frequently fail to solve large problems within a reasonable timeframe.

Previous studies have already connected several elements that were discussed in this review. Table 2.6 compares the studies that come closest to quantifying how uncertainty, robustness and cost relate to each other. Only Yap and Van Oort (2018) quantify the trade-off between an efficient and a robust schedule in terms of disruption cost for passengers and operators, but that is focused on driver schedules and on a single disruption. The other studies show how realistic EC, operational uncertainty and the cost of robustness can be accounted for in the planning of charging infrastructure, charging strategies and electric bus scheduling, but quantify the trade-off only partly. No study was found that connects them for the vehicle schedule of electric buses. The literature therefore leaves three gaps. First, the accuracy of EC models is not assessed on the level at which schedules are used. Second, EC uncertainty is not combined with scenarios and a simulation of a whole vehicle schedule. Third, little attention has been paid to how the EC value that is adopted during tactical electric bus scheduling affects both the robustness and the efficiency of the resulting schedules, when evaluated under a common set of EC scenarios and valued from the perspective of both passengers and the operator. The last row of table 2.6 shows how this study positions itself.

Table 2.6: Comparison of the studies that come closest to quantifying the trade-off between robustness and efficiency

Study	Planning problem	Uncertainty	How robustness and cost are assessed	Trade-off quantified
Yap & Van Oort (2018)	Driver scheduling	Large, non-recurrent disruption	Passenger and operator disruption cost against driver schedule cost	Yes, in cost
Momen et al. (2025)	Charging infrastructure design	EC variability, sparse data	Share of infeasible scenarios against infrastructure cost	Yes, in cost
Tang et al. (2019)	Electric bus scheduling	Stochastic traffic (trip time)	Breakdowns and delay cost; buffer distance	Partly
Bie et al. (2021)	Electric bus scheduling	Travel time and EC	Stochastic optimization of delays, EC and cost	Partly
De Bruin et al. (2023)	Electric bus scheduling	Driving times	Punctuality and cost via DES and Monte Carlo	Partly
Manzolli et al. (2025)	Fleet charging management	EC	Robust versus deterministic charging cost	Partly
This study	Electric bus scheduling	EC	Trip completion, passenger delay cost, operator cost and net change in cost via DES and Monte Carlo	Yes, in cost

Taking all of the above into account, the following scientific contributions are identified:

- the development and validation of a data-driven EC model, based on a full year of real trips of the Dutch region IJssel-Vecht, that quantifies the influence of external factors and is validated both on trip and on vehicle task level;
- a scenario-based simulation framework that combines the EC model, scenarios for temperature and passenger numbers and the unexplained variation of the EC model (Monte Carlo) to assess the robustness of electric bus schedules under EC uncertainty;
- a quantification of the trade-off between robustness and efficiency, by comparing schedules that are generated with different EC planning values under the same scenarios and expressing the outcomes in reliability, passenger delay and operator cost.

This study positions itself on the three gaps with a combination of methods that have been used separately before: a data-driven EC model validated on vehicle task level, a scenario-based simulation with Monte Carlo runs, and a valuation of the consequences for passengers and the operator. The next chapter describes how this is done.

3

Methodology

This chapter presents the methodology of the study. It builds on the literature review in two ways. The review showed which factors influence the EC of electric buses and which modelling techniques are suitable to describe them, and it showed that the effect of the EC value that is assumed in the planning of vehicle schedules on robustness and efficiency has hardly been quantified. The methodology is described in general terms, so that it can be applied to any region for which the required data is available. The application to a specific case is introduced in chapter 4.

The study does not develop an optimisation model for the vehicle schedules. Instead, it assesses the consequences of a choice that is made within the optimisation, the EC value that is used as input, for the performance of the schedule during operations, and it values robustness and efficiency in a way that allows for a comparison. The main research question to be answered is therefore: *How does the energy consumption value that is assumed in the scheduling of electric buses shape the trade-off between robustness and efficiency, for public transport facing energy consumption uncertainty?*

To accommodate the main research question, a set of four sub-questions is formulated:

1. *What external factors influence energy consumption of battery electric buses in a Dutch public transport network and what is the impact on energy consumption for each factor?*
2. *What scenarios can be defined to test the robustness of a schedule for a battery electric bus fleet in a Dutch context?*
3. *What is the estimated robustness of an actual schedule when values of external factors change according to the given scenarios?*
4. *How does a change in the energy consumption input rate for fleet scheduling influence the robustness and efficiency of the given output?*

The first sub-question dives into the energy consumption of the buses and is focussed on statistical analysis. The second sub-question will use that information to find relevant scenarios, which, again, can be used to assess robustness in the third sub-question. There, simulations are run to assess the schedule performance in terms of reliability when the values of important external factors change. The fourth sub-question aims to explore how lower or higher energy consumption rates during optimization influence robustness. The corresponding schedules can be used to derive statistics on efficiency. The complete approach is visually represented in figure 3.1. The methods are further elaborated in the remainder of this chapter.

The sub-questions are connected through the output of the previous step. The EC model of SRQ1 provides, for every bus type, a function that estimates the EC per kilometre of a trip from the external factors. It also provides the unexplained variation, in the form of the residuals on vehicle task level. The scenarios of SRQ2 define the values of the most influential factors (temperature and pax) together with their probability. In the simulation of SRQ3 the EC model is used to calculate the EC of every

trip of a schedule under the conditions of a scenario. In the sensitivity analysis of SRQ4 the simulation is repeated many times, with a random error on the EC that is based on the residuals of the EC model (Monte Carlo simulation), for every scenario and for each of the schedules. The outcomes per scenario are then weighted with the probability of the scenario to arrive at expected annual outcomes per schedule, which are used to quantify the trade-off between robustness and efficiency. The general methodology is applied to a specific case, which is introduced in chapter 4. Chapter 5 describes the data for that case, after which the methodology is carried out in chapter 6.

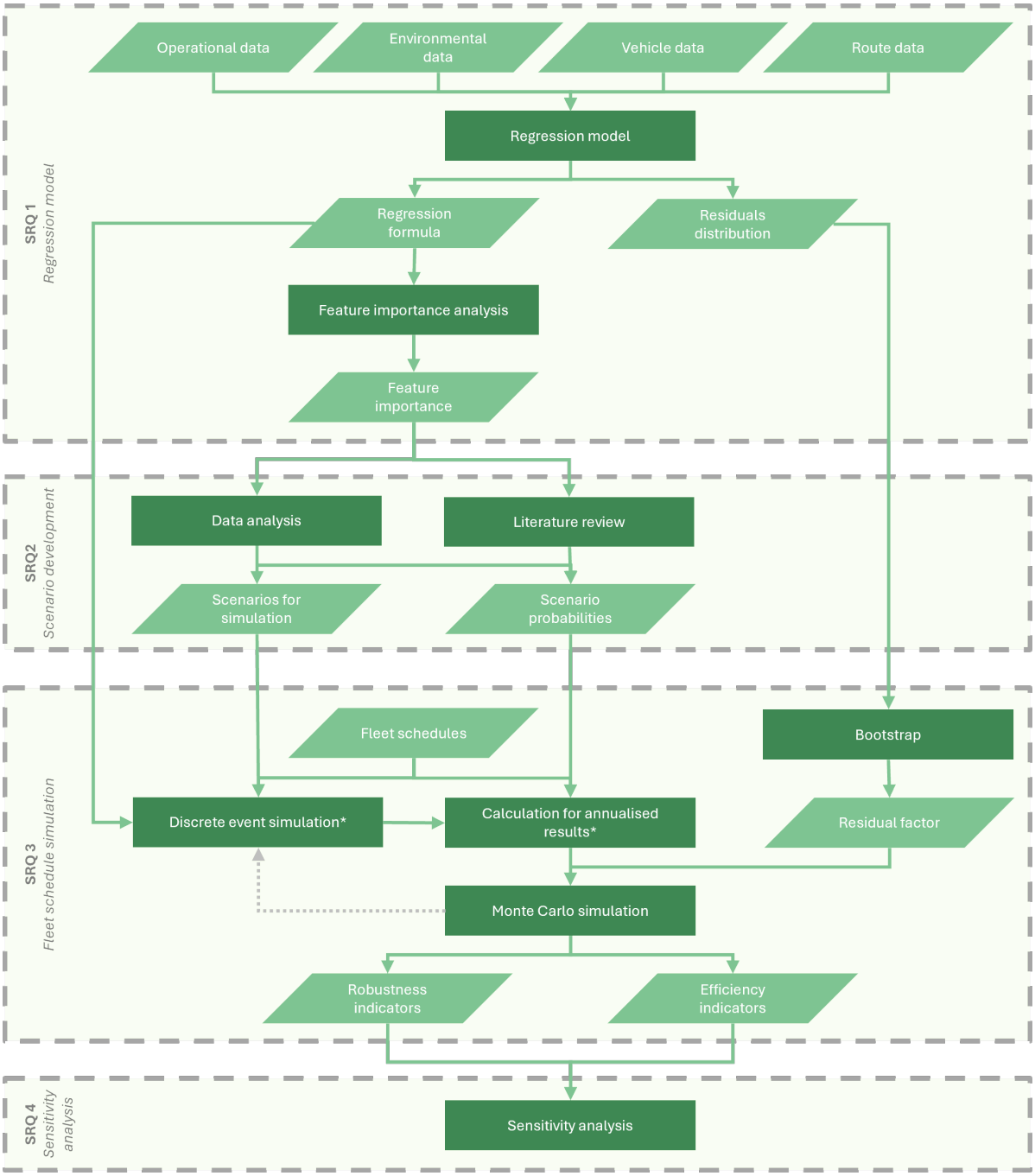


Figure 3.1: Schematic overview of the research approach

3.1. Regression model for energy consumption estimation

SRQ1 is key in understanding which, and, in what way, external factors – everything besides the technical aspects of the bus itself – influence the energy consumption of the buses.

To start, the knowledge from the literature review is combined with the data that operators typically collect to find which factors should be taken into account in the analysis. The input data comes from two sources: the operator, through the sensors on the buses and the planning systems, and a national weather institute, which provides open weather data. An overview of the input data is provided in table 3.1. After cleaning, the data is assigned to a specific trip. This is a specific route segment with an origin and destination stop and all stops in between, according to a predetermined schedule. A notable omission compared to the reviewed literature is the impact of slope or height difference, for which data is not available in the standard data of an operator.

Table 3.1: Data used in the EC analysis

Vehicle data	**Type**	**Unit**
bus type	categorical	
battery capacity	integer	kWh
energy consumption	integer	$\mathrm{kWh}_{/\mathrm{km}}$
initial state of charge	integer	%
state of health	integer	%
Operational data		
average speed	continuous	$\mathrm{km}_{/\mathrm{h}}$
difference travel time	time	seconds
Route data		
route type	binary	
distance travelled	continuous	km
pax	integer	
Environmental data		
temperature outside	continuous	$^{\circ}$C
precipitation	integer	
headwind	continuous	$\mathrm{m}_{/\mathrm{s}}$

Most factors follow directly from the sensor data, but some have to be computed. The EC of a trip is the difference between the SoC at the start and at the end of the trip, multiplied by the usable battery capacity (the nominal capacity corrected for the state of health), and divided by the trip distance to arrive at the EC per kilometre. The average speed is the distance of the trip divided by its realised travel time, and the difference in travel time is the realised minus the scheduled travel time. Pax is the number of check-ins on the trip. The headwind is the component of the wind speed in the direction of travel, which follows from the wind speed, the wind direction and the bearing between the origin and the destination stop.

For SRQ1 the modelling approach is based on the outcomes of the literature review, where the advantages and disadvantages of each method are described. Multivariate non-linear regression (MNR) can provide a solution to the unrealistic linearity assumption for temperature, while still providing insights into the influence of each factor on EC. The MNR uses independent variables to explain the variance between measurements of the dependent variable. In this study the EC factors are the former and the EC per kilometre is the latter. Using EC per kilometre, instead of EC per trip, prevents the trip distance from explaining almost all of the variance (see chapter 5). An ordinary least squares (OLS) type MNR is trained, separately for each bus type, with five-fold cross-validation (CV). OLS is the most conventional type of regression, fast to compute, and, when used for a large data set of trips, has no problems with over-fitting. For smaller sample sizes, one could look to ridge or lasso regression, as those prevent over-fitting by penalising features that contribute marginally to the model's performance. The CV is used to prevent that the performance is judged on a biased selection of trips. The MNR is set to degree 2, which implies that both linear and quadratic relationships are taken into account. Degree 2 also multiplies every factor with all other factors, and these combinations (interaction effects) also feature in the model.

The model for the EC per kilometre y_i of trip i, with the nine factors x_{ij} (standardised), is

$$y_i = \beta_0 + \sum_j \beta_j x_{ij} + \sum_{j \leq k} \beta_{jk}\, x_{ij} x_{ik} + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{3.1}$$

where β_0 is the intercept, β_j the coefficients of the linear and squared terms (the case $j = k$), β_{jk} the coefficients of the interaction terms and ε_i the unexplained part of the EC. The squared terms capture curved relationships, such as an increase in EC at both low and high temperatures, and the cross terms capture interactions between factors. Writing the model as $y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i$, the coefficients are estimated with ordinary least squares:

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left(y_i - f(\mathbf{x}_i, \boldsymbol{\beta})\right)^2. \tag{3.2}$$

In other words, the coefficients are chosen such that the sum of the squared differences between the measured and the predicted EC over all trips is as small as possible. As the model is linear in its coefficients, this has an exact solution that is fast to compute.

The aim is to create an EC model that can be broadly applied to various electric bus types and in various regions. Performance is measured in explained variance with the R^2, and in the root mean square error (RMSE) and mean absolute error (MAE) as indicators for the accuracy (see section 3.1.2). The higher the R^2, the more variance can be explained through the model. For RMSE and MAE, on the other hand, the number should be as low as possible. One problem for a broadly applicable model is the omission of terrain height or slope in this data set. In most public transport regions of the Netherlands, the slope has minor influence as the country is relatively flat. For other regions, however, the height difference does play an important role and this method might come up short in such cases.

In general, the data for most of the factors should be available for all public transport operators. The factors that are included in the model follow from the literature review and from the data that is available. Because the data set is large, all nine available factors are included in the model, and no factors are excluded on the basis of a threshold for their individual contribution. Instead, the importance of every factor is assessed afterwards, so that the factors that are relevant for the scenarios can be selected on the basis of the results. The importance is assessed in three ways. First, the partial dependence plots (PDP) show how EC changes when a single factor changes and all others are kept at their mean value. The range of EC within the 5$^{\text{th}}$ to 95$^{\text{th}}$ percentile of the factor is used as an indicator for the size of the effect. Second, the drop in R^2 is determined when the values of a factor are randomly shuffled (permutation importance). Third, the drop in R^2 is determined when a factor is left out of the model and the model is fitted again (drop-column importance). The latter is the most robust indicator for the added value of a factor, as the effect of correlated factors is taken over by the remaining factors. Because correlated factors can eat away each other's effect, the correlation between all factors is also assessed, with the Pearson correlation coefficient. A coefficient of 0.7 or higher is generally accepted as strong. The influence of the interaction effects is determined by adding every possible interaction, one at a time, to a model that only contains the main and squared terms, and comparing the held-out R^2.

3.1.1. Speed dynamics indicator

Most factors included in the EC modelling are directly related to the input data, with one exception: speed data. The CANbus system of the bus registers speed up to multiple times per second, which gives a speed profile for every trip. From such a profile, indicators can be derived that describe how smoothly or aggressively a bus is driven, such as the positive kinetic energy, the relative positive acceleration, the number of hard acceleration and braking events and the jerk. Combined, these form a speed dynamics score. An exploratory analysis was carried out to see whether such a score improves the estimation of EC. The data requirements of the score proved too high to apply it to a complete year of trips, so the score is not part of the EC model. The indicators and the analysis are described in appendix B.

3.1.2. Validation of model performance

The EC model is validated on two levels. First, the performance is assessed on trip level, for which the model is trained with five-fold cross-validation. The data is split randomly into five parts, and each

part is predicted by a model that is trained on the other four. In this way every trip has an out-of-fold prediction, so the performance is not overstated by predicting trips that were used for training. The performance is expressed in:

- the explained variance (R^2);
- the root mean square error (RMSE) and mean absolute error (MAE), in kWh/km, to express the size of the error in the unit of the model;
- the 5th and 95th percentile of the residuals, to express the spread of the error, including whether the model tends to over- or underestimate.

The MAPE, as described in the literature review, is not used as the main indicator, because trips with an EC of zero cannot be included, and because the discrete steps of the SoC make the percentage error very large for short trips.

As the EC is measured as an integer percentage of the battery capacity, short trips suffer from large rounding errors (see chapter 5). The second level of validation is therefore the vehicle task level. Here, the EC predictions of all trips driven by a vehicle on a single day are combined, and compared to the measured EC of that day. The same indicators are used. As the simulation uses a simplified version of the EC model (see chapter 6), the validation on vehicle task level is repeated with these simplifications. The residuals at this level are also used as the random error in the Monte Carlo simulation of the sensitivity analysis. In short, the first step of the methodology results in an EC model with known accuracy and a distribution of its unexplained variance, which are both used in the steps that follow.

3.2. Scenario development

SRQ2 is the first step in the robustness assessment. In order to quantify how robust a schedule is, it is necessary to define the range of conditions that the schedule has to operate in. For this, the most influential factors from the previous question are further explored. Knowing how large the impact of a certain factor on EC is, the next step is to find more information about the occurrence of values that result in high EC values. This can be partly derived from the datasets that are used for the EC model described in the previous section. However, using only historical data of a given period, may lead to biased estimations for future scenarios. That is why additional literature will be consulted to check whether historical patterns within the data set, align with expected future trends. The aim of this question is to generate a set of scenarios $\mathcal{S}$ that give insight into how the schedule performs under both extreme scenarios and more common, but still impactful scenarios. Each scenario will have a weight based on the probability of its occurrence. The weight is later used to transform scenario-based outcomes to expected schedule-based outcomes. This is explained in more detail in section 3.3.3.

3.3. Simulation of the fleet schedule

SRQ3 seeks to quantify the performance of a schedule $p \in \mathcal{P}$ for every scenario $s \in \mathcal{S}$. This is done by using SimPy, a python package for discrete event simulation. The simulation provides a way to evaluate the effects of each scenario on reliability and other indicators, and ultimately describe the robustness of the given schedule. In this experimental set-up, only the external factors according to the scenario are changed, while the other parameters stay the same. A flowchart of the algorithm is shown in figure 3.2. The simulation starts by loading the configuration of the fleet and the schedule and assigning each vehicle task to a bus. Next, the scenario and parameter levels are set and a simulation run is initiated. As the DES works with discrete time steps - in this case one minute - all vehicle tasks are run in parallel. The events are either trips or periods of charging. Whenever an event starts or ends, the status of a bus changes. There are three statuses: performing trip, waiting for the start of the next trip (i.e. sitting idle) or charging. At the start of the run, it is assumed that all buses have a SoC of 100%.

3.3.1. Discrete event simulation

The calculation of the EC is based on the regression model from SRQ1. The model uses the values of the external factors that are defined by the scenario (temperature and pax), the SoC at the start of the trip (which follows from the simulation itself) and the average speed that follows from the schedule.

For the other external factors, the mean values from historical data are used. The random error of the model is not included in the discrete simulation, so the calculation is deterministic. The uncertainty is purely the variation in the discrete calculations between all the scenarios. In the sensitivity analysis the random error is added to the EC through a Monte Carlo simulation (see section 3.4). A trip is only performed when two conditions hold: the expected SoC at the end of the trip is not lower than the minimum SoC, and the bus can still reach a charger from the destination of the trip without dropping below that minimum. If the SoC is insufficient to perform a trip, the planned trip is cancelled and a new dead-heading trip to a charger is performed. The algorithm thereby favours fast chargers before slow chargers. While dead-heading trips are generally more direct than scheduled trips between the same origin and destination points, not all origin-destination (OD) pairs might feature in the data from the schedule. Instead, for unplanned dead-heading trips the Dijkstra algorithm searches for the shortest path between all possible combinations of existing trips for the same OD pair. The minimum SoC is a safety margin below which a bus is not allowed to drop, and is set at 5%.

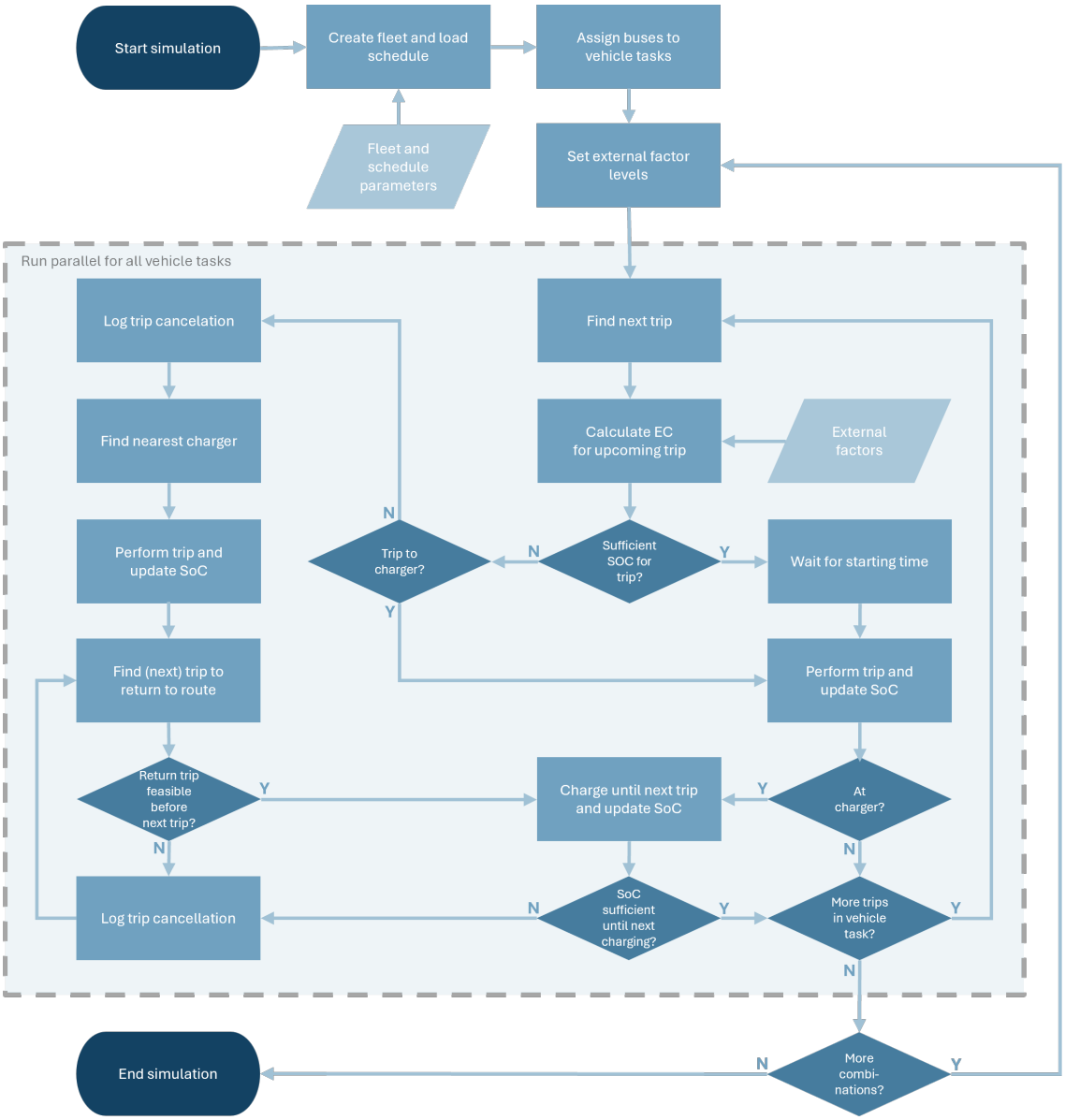


Figure 3.2: Flowchart of the simulation model used to assess robustness

If the SoC is sufficient, the bus waits for the starting time and the trip is performed. When the trip ends at a charger, the bus automatically initiates charging until the next trip starts. Charging is linear at the

nominal power of the charger. Should all chargers be occupied, the bus joins a queue based on first in, first out. A bus that is fully charged releases its charger when another bus needs it, which takes a switching time of five minutes. If the SoC at the start of a trip is not sufficient to fulfil all trips until the next charging moment, the charging continues and the trip is cancelled. At the same time, a check is done for a possible return trip. If it is possible to get to the next starting location before the start of the next trip, and if there is enough SoC charged for all the trips until the next charging moment, the charging lasts until the return trip. Otherwise the trip is cancelled and there is a check whether it is feasible to reach the trip thereafter.

For the simulation to be reproduced, the following input is required: the schedule (the service trips with their origin, destination, start time and distance, assigned to vehicle tasks, including the planned deadheading trips and charging events), the bus type of every vehicle task with its usable battery capacity, the location, power and number of chargers, and the EC model. The time step of the simulation is one minute, and all buses start the day with a full battery. Charging is linear at the nominal power of the charger, and a bus that is fully charged releases its charger for another bus after a switching time of five minutes. The minimum SoC is 5%, and the speed of a trip is the scheduled distance divided by the scheduled travel time.

3.3.2. Monte Carlo simulation

The discrete calculation of the EC ignores the unexplained variance of the EC model. That is why a Monte Carlo simulation is introduced, in which the EC of the trips of a vehicle task is multiplied by a factor that corresponds to a random residual. This factor is obtained by a bootstrap from the observed ratios between measured and predicted EC of vehicle tasks, specified per depot and per vehicle type. A vehicle task keeps its factor in all scenarios of a run. The Monte Carlo simulation consists of 1,000 runs for each scenario, for each schedule and for each depot. After all the runs are completed, the average value over all runs is taken for every indicator, and that is considered the outcome. The 5^{th} and 95^{th} percentile of the runs indicate the range of the outcomes.

After all the trips are simulated, calculations can be made per scenario on several aspects of the schedule's performance. These calculations are discussed below.

3.3.3. Indicators

To make a quantitative judgement on how robust and efficient a schedule is, transparent indicators are needed. The indicators follow from the logs of the discrete event simulation, in which all cancelled trips are recorded, and from the schedule itself. They are divided into robustness and efficiency indicators. The former focus on the effect of trip cancellations (due to insufficient battery charge) on passengers, in the form of the share of completed trips and societal cost. The latter describe the efficiency that is directly derived from the schedule and the cost that the operator has. The selection follows the literature on robustness (chapter 2): the consequences of disturbances are valued for passengers, as in Yap (2014), and set against the cost of the operator, and the efficiency is expressed in the resources that are used for service. Since many concessions pay the operator per kilometre or hour of service, the share of the time and distance that is used for service is also an indication of how many resources are spent on the activity that earns the money.

The results are obtained on four distinct levels, of which the hierarchy is depicted in figure 3.3. The outcomes start on trip level, where the impact of a cancelled trip is given by the impact for the passengers and the expected number of passengers on that trip. The outcomes of all cancelled trips in the same scenario under the same schedule are summed to form the outcomes at scenario level. Then, the outcomes of that scenario are multiplied by the expected number of days of that scenario to form annualised results. This is done for all scenarios under the same schedule and their results are again summed, which creates the outcomes on schedule level. On schedule level additional indicators can be derived directly from the schedule, mainly on efficiency. Combining all schedules of the same depot then creates the sensitivity analysis, which is explained in the next section. The essence of an indicator stays the same across the levels, only the scale changes.

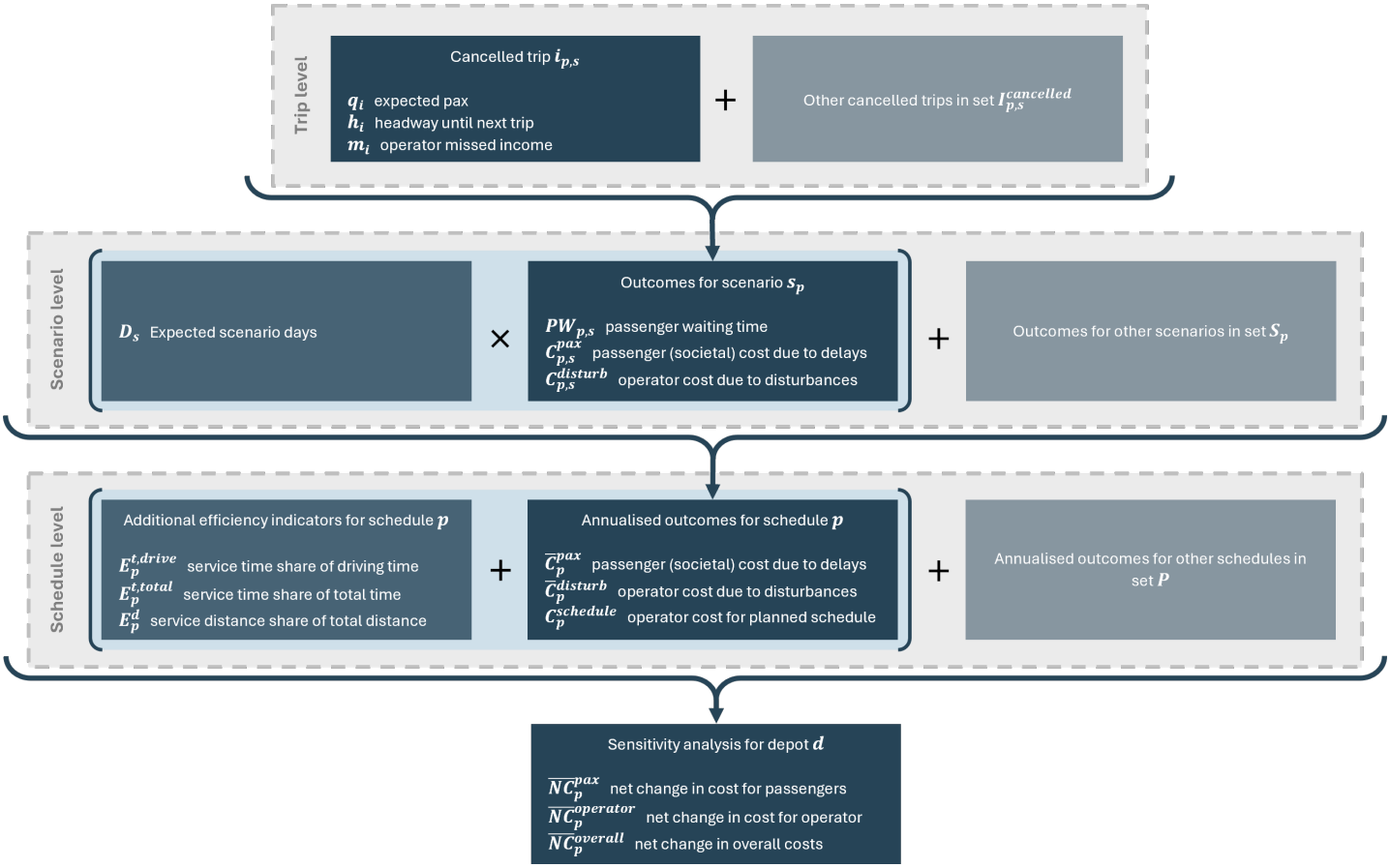


Figure 3.3: Overview on the indicators on the four distinct levels

Robustness indicators

The first indicator is not shown in figure 3.3, but can be calculated in a relatively easy manner.

From a passengers' perspective, a trip is successful if they can board their trip of choice. When a trip is cancelled, this has a negative impact on their satisfaction. Having a high reliability for trip completion is therefore desirable. **Trip completion rate** (for the schedule $p \in \mathcal{P}$, according to scenario $s \in \mathcal{S}$) is expressed as the percentage of completed service trips compared to the total amount of service trips over the course of one day.

$$R_{p,s} = \frac{N_{p,s}^{\text{comp}}}{N^{\text{serv}}} \tag{3.3}$$

Where $N_{p,s}^{\text{comp}}$ is the amount of completed service trips for schedule p in scenario s and N^{serv} is the total amount of scheduled service trips.

The impact of cancelled trips on passengers is also expressed as the total delay that passengers experience, in hours and in societal cost. The delay of a passenger on a cancelled trip is the time until the next viable trip, so the headway. The calculation is first made for every cancelled trip, and then summed over the trips of a scenario. At schedule level, the scenario results are multiplied by the expected number of days of the scenario and summed. There are two simplifications for this calculation:

- The amount of delay is assumed to be the headway between trips from the same line and in the same direction. When a line overlaps with another line for a significant share of the line (more than $\frac{2}{3}$ of the stops), and has a complementing headway (i.e. halves it), the headway is multiplied by a factor 0.6. For small overlap (between $\frac{1}{3}$ and $\frac{2}{3}$ of the stops) this factor is assumed to be 0.8. The reasoning behind this is that a portion of pax will be able to reach their destination with the bus from another line and thus do not have to wait the full headway time for the next connection.

- Missed transfers or other forms of delay propagation are not considered.

For the **passenger delay** in hours (for the schedule $p \in \mathcal{P}$, according to scenario $s \in \mathcal{S}$), the following equation holds:

$$PW_{p,s} = \sum_{i \in \mathcal{I}_{p,s}^{\text{cancelled}}} q_i h_i \tag{3.4}$$

Where $\mathcal{I}_{p,s}^{\text{cancelled}} \subseteq \mathcal{I}$ is the subset of cancelled service trips within the set of all service trips $\mathcal{I}$, q_i is the amount of pax on trip i and h_i is the amount of headway between trip i and the next viable trip.

For the **passenger delay cost** (for the schedule $p \in \mathcal{P}$, according to scenario $s \in \mathcal{S}$), the experienced delay time for passengers on a cancelled trip is multiplied by a factor for extra out-of-vehicle waiting time and a factor for delayed arrival time. The pax per trip is derived from historical check-in data.

$$C_{p,s}^{passenger} = PW_{p,s} \times VoT^{IV} \left(\lambda^{\text{wait}} + \lambda^{\text{arr}}\right) \tag{3.5}$$

Where VoT^{IV} is the in-vehicle VoT, λ^{wait} is the factor for out-of-vehicle waiting time and λ^{arr} is the factor for unexpected arrival delay. They are added together because the cancellation of a bus leads to discomfort from both having to wait longer outside and arriving later at the destination. Both are for the duration of the headway until the next bus.

While low robustness also affects the operator, this effect is incorporated in the efficiency indicators in the form of the cost of cancelled trips, which are introduced next.

Efficiency indicators

The last two indicators that are calculated on scenario level, have to do with the **cost associated with disturbances for the public transport operator** due to energy consumption uncertainty (for the schedule $p \in \mathcal{P}$, according to scenario $s \in \mathcal{S}$). These can come in the form of missed income and - when a threshold value is exceeded - fines. These fines tend to differ between public transport authorities and are therefore specified with the introduction of the case study. The same holds for the missed income. In most Dutch public transport areas, the public transport operator is paid per hour of passenger service in the form of subsidy, alongside the revenue from ticket sales. Cancelling a trip results in the cancellation of the payment from the public transport authority for that trip, which is considered missed income. The ticket revenue can be regained if the passengers choose to wait for the next bus. This study assumes that they do so. At the same time, cost components - with the exception of variable cost per km - will not change. The driver, for example, still has to be paid while sitting idle. That is why, from a financial perspective, the public transport operator should aim to reduce trip cancellation.

$$C_{p,s}^{\text{disturbance}} = \sum_{i \in \mathcal{I}_{p,s}^{cancelled}} \left(m_i - c_i^{\text{saved}}\right) + c^{fine}\, N_{p,s}^{\text{canc}} \tag{3.6}$$

Where m_i is the missed income from trip i (the service hours of the trip multiplied by the subsidy per service hour), c_i^{saved} is the saved variable cost from trip i, c^{fine} is the fine per cancelled trip and $N_{p,s}^{\text{canc}} = |\mathcal{I}_{p,s}^{cancelled}|$ is the number of cancelled trips. In many concessions the fine depends on the share of cancelled trips, but a fixed amount per trip is used here so that the depots in the case study can be compared (see chapter 4).

For this study, efficiency is expressed through the cost associated with the schedule, the number of buses required to operate the schedule, and the share of time and distance that is used for service trips compared to the total. As the service trips are the same for every schedule, the differences between the schedules are found in the dead-heading trips, the charging events and the waiting time. The planned schedule cost consists of a **fixed cost** per bus, which is charged for every day of the year because the cost of a bus does not go away when it is not used, and a **variable cost** for the distance that is driven (service and dead-heading) on the days that the schedule is operated.

$$C_p^{schedule} = c^{fixed} \times v_p \times D^{\text{cal}} + c^{variable} \times d_p^{total} \times D^{\text{op}} \tag{3.7}$$

Where c^{fixed} is the fixed cost per bus per day, v_p the number of buses (vehicle tasks) required for schedule p, D^{cal} the number of calendar days in a year, $c^{variable}$ the variable cost per km, d_p^{total} the total distance driven per day under schedule p and D^{op} the number of operating days. In practice, only the difference between the schedule cost of schedule p and the benchmark schedule, $\Delta C_p^{\text{schedule}}$, is of interest, as it determines what has to be paid for a change in robustness.

The **efficiency expressed in time and distance** is as follows:

$$E_p^{t,\text{drive}} = \frac{t_p^{service}}{t_p^{driving}}, \qquad E_p^{t,\text{total}} = \frac{t_p^{service}}{t_p^{total}}, \qquad E_p^{d} = \frac{d_p^{service}}{d_p^{total}} \tag{3.8}$$

Where $t_p^{service}$ is the number of hours spent on service trips, $t_p^{driving}$ the hours spent on service trips and dead-heading trips and t_p^{total} the total number of hours of the vehicle tasks, which also includes charging and waiting. The distance efficiency E_p^d is calculated in the same way, but then for distance, where $d_p^{service}$ and d_p^{total} are the service distance and the total distance driven. Service time is the time spent on timetabled trips, driving time is service time combined with the time spent on dead-heading trips and total time is time that all vehicles tasks spent outside of the depot. The difference between service, driving and total time (and between service and total distance) matters, because it is an indication of how many of the resources are spent on service. In most cases, operators are only paid for the service that is offered, while the buses and drivers are paid for all hours and kilometres. Finally, the charging pattern is described by the average duration of a charging event, the number of charging requests per year and the number of requests that find all chargers occupied.

Annualised results

The first step of the sensitivity analysis is to assess the performance of the benchmark schedule. This can be either an existing schedule used by the operator, or an experimental one. The schedule is simulated under all scenarios. The outcomes of a scenario are multiplied with the expected number of days per year on which the scenario occurs, D_s. This number follows from the number of operating days of the schedule D^{op} (the days on which the normal weekday schedule is operated), the probability of the temperature of the scenario π_s^{temp} and the weight of the pax level π_s^{pax}:

$$D_s = D^{\text{op}} \cdot \pi_s^{\text{temp}} \cdot \pi_s^{\text{pax}}, \qquad w_s = \frac{D_s}{\sum_{s' \in \mathcal{S}} D_{s'}} \tag{3.9}$$

where w_s is the corresponding weight of the scenario, that is used for averages. For example, if scenario A is expected to occur on 10% of the days that the schedule is in use in a year, then the outcomes of that scenario are multiplied by that number of days. Summing over all scenarios leads to an expected annual number for different indicators. Note that these annual numbers only concern the scenarios that are included in the analysis. They are not the total annual outcomes of a schedule, as the days on which none of the scenarios apply (for example warmer days) are not simulated, and are expected to lead to no cancellations. Additionally, there are indicators related to the schedule that do not fluctuate depending on the outcomes of a scenario. An example is the planned dead-heading distance.

3.4. Sensitivity analysis

The final step of the methodology answers how the EC value that is assumed in planning shapes the trade-off. In this section, schedules that were planned with different EC values are generated, they are assessed with the simulation of the previous section, and their outcomes are compared in terms of robustness and efficiency.

After the benchmark schedule is assessed, **SRQ4** aims to perform a sensitivity analysis with new schedules where the EC as input for the optimization tool is increased or decreased by 10% or 20% compared to the input levels of the benchmark schedule. The steps of 10% are chosen, because they represent a balance between a high and a realistic change. A change of 50% for example, would lead to interesting changes in the schedule's performance, but the chance that such a change is necessary is low.

The values describe the EC that is assumed in the optimisation. A value of +20% means a higher assumed EC, implying that the buses in the optimisation can cover less distance on a full charge, as

the range is the usable battery energy divided by the EC per kilometre. A higher EC input generally leads to a more robust schedule, as there is more energy redundancy on days with a high EC, and vice versa. At the same time, a higher input generally leads to a less efficient schedule, as the lower implied range requires a bus to make more trips to the charger (covering extra distance and taking extra time) and to sit idle at the charger for longer. Note that the impact depends strongly on the EC rate that was used for the benchmark schedule.

3.4.1. Fleet schedule optimization

The preferred method - designing an optimization model - is not possible due to time constraints. Instead, optimization software that the public transport operator uses for its own operations provides the fleet schedules that will be analysed further. The main objective of the optimization tool is to minimize an internal cost indicator. This consists of both financial cost and cost penalties for non-financial constraints. The problem that is solved is the electric vehicle scheduling and charging problem: the service trips have to be assigned to vehicle tasks (buses), and dead-heading trips and charging events are planned in between. The main sets are the service trips, the buses and the chargers. The main decisions are which bus performs which trip, and when and where a bus charges. The goal is to minimise the cost of the fleet schedule, such that no bus runs out of energy. The optimization has the following assumptions:

- There is a flat rate for EC per bus type. This means that for all trips service by the same type, a single rate is used, rather than a rate based on route or operational characteristics.
- Charging follows a linear curve.
- Partial charging is allowed.
- The number of available chargers is limited.

3.4.2. Quantification of the trade-off

The final step is to define the actual trade-off between robustness and efficiency between the different schedules. The idea of a more robust schedule is to save cost when operating under disruptions. It is therefore interesting to see whether more robust schedules indeed lead to cost savings under scenarios with a higher EC, and whether less robust schedules lead to higher cost. This is evaluated by setting the change in schedule cost against the change in the cost of trip cancellations, for the passengers and for the operator. The scenario outcomes of each schedule are multiplied by the expected number of days that the scenario occurs (D_s), which results in the expected annual cost of cancelled trips for the operator and for the passengers:

$$\overline{C}_p^{\,\text{disturbance}} = \sum_{s \in \mathcal{S}} D_s \, C_{p,s}^{\text{disturbance}}, \qquad \overline{C}_p^{\,\text{passenger}} = \sum_{s \in \mathcal{S}} D_s \, C_{p,s}^{\text{passenger}} \tag{3.10}$$

The schedule cost is not multiplied by the scenario days, as it is already an annual number, and the fixed cost is charged for all days of the year, as explained above.

The net change in cost of schedule p compared to the benchmark schedule p_0 is expressed for the passengers, the operator and overall:

$$\Delta C_p^{\text{pass}} = \overline{C}_p^{\,\text{passenger}} - \overline{C}_{p_0}^{\,\text{passenger}} \tag{3.11}$$

$$\Delta C_p^{\text{op}} = \left(C_p^{\text{schedule}} - C_{p_0}^{\text{schedule}}\right) + \left(\overline{C}_p^{\,\text{disturbance}} - \overline{C}_{p_0}^{\,\text{disturbance}}\right) \tag{3.12}$$

$$\Delta C_p^{\text{all}} = \Delta C_p^{\text{pass}} + \Delta C_p^{\text{op}} \tag{3.13}$$

In other words: how much additional (societal) cost was incurred, or averted, and what was the price paid, or the saving made, for this schedule adjustment. A negative value is a saving, which means that the schedule pays for itself. The schedule with the lowest overall net change in cost is the schedule with the optimal robustness for the case study that is under investigation. The change in cost is expressed relative to the annual operator cost of the benchmark schedule, $C_{p_0}^{\text{schedule}} + \overline{C}_{p_0}^{\,\text{disturbance}}$, being the sum of the planned schedule cost and the expected cancellation cost of that schedule. This relative number makes it possible to compare depots of a different size. It does not include other costs of an operator, such as personnel and overhead. If the schedule that is used as benchmark was generated with

different settings than the alternative schedules, the planned schedule cost of the benchmark is not a fair reference. In that case, an approximation of the planned schedule cost without this effect can be used instead. This is applied in the case study and explained in section 6.4.1.

3.4.3. Additional analysis on the impact of crew scheduling

The cost of personnel is not part of the schedule cost in the main analysis. To illustrate the possible effect on the driver schedules, an estimate is added in an extended analysis. All driving time (service and dead-heading trips) is valued at a cost per daily driving hour $c^{\text{pers,drive}}$, and the idle hours in the vehicle tasks (charging and waiting) at a lower cost per hour $c^{\text{pers,idle}}$:

$$C_p^{\text{personnel}} = c^{\text{pers,drive}}\, t_p^{driving} + c^{\text{pers,idle}} \left(t_p^{total} - t_p^{driving}\right) \tag{3.14}$$

This estimate is added to the schedule cost C_p^{schedule}, so that the differences between the schedules in dead-heading and idle time become visible in the net change in cost. The idle time is valued lower, as not all of that time has to be paid time, but more of it makes it harder to build efficient driver schedules. The values of both parameters are given in chapter 4, and are varied in a sensitivity analysis in the results, together with the fine for a cancelled trip.

3.5. Conclusions on methodology

The methodology consists of four connected steps. A regression model estimates the EC per trip, and its validation on vehicle task level determines how accurate it is where it matters. Scenarios, based on the most influential factors and their probability, define the conditions under which a schedule is tested. A discrete event simulation with Monte Carlo runs translates a schedule and a scenario into cancelled trips, from which robustness and efficiency indicators are derived. Finally, schedules that are planned with different EC values are compared under the same scenarios, which quantifies the trade-off in terms of cost for passengers and the operator.

The methodology has several limitations. The regression model lacks the road gradient, and the simulation uses simple decision rules and a single random factor per vehicle task. The alternative schedules rely on optimisation software of which the exact workings are not known, and the cost of personnel can only be estimated without a crew scheduling step. These limitations are taken into account in the interpretation of the results. The next chapter introduces the case study to which the methodology is applied.

4

Case study

The methodology described in the previous chapter is general by design: it can be applied to any transport region for which energy consumption data, a vehicle schedule and a set of cost parameters are available. To test the methodology, it is applied to a real case. This chapter introduces that case. First, the choice for the IJssel-Vecht concession is explained, after which the two depots that are studied in detail are introduced. Then the schedules that are compared in the sensitivity analysis are described, followed by the parameters that are used to translate the outcomes of the simulation into costs. The chapter closes with a reflection on how far the outcomes of this case study can be generalised.

4.1. The IJssel-Vecht concession

This research is set out by EBS, so the case study is a transport region in which they operate. At the end of 2025 EBS held contracts for four regions in the Netherlands, three of which are operated with battery electric buses. Historical data from the whole year of 2025 is considered. IJssel-Vecht was the preferred region, because it provides the largest data set on EC for different bus types. This is important for the regression model, as more data points improve the reliability of the estimations, especially for a polynomial model.

The IJssel-Vecht region is one of the largest in the Netherlands in terms of area, but includes only a handful of large cities. Apeldoorn, Deventer, Lelystad and Zwolle have their own set of urban lines, whereas the network of Harderwijk and Kampen is limited to a single intra-city line. The remaining part of the network consists of regional lines that connect these cities and the villages in between.

EBS operates four electric bus types in this region, all part of the BYD K-series: the 9m bus, the 12m bus, the 12m pantograph bus and the 13m bus. EBS splits the 12m bus into two variants for marketing purposes. One has a normal livery, while the other features the livery of the express line (SnelRRReis). The technical aspects are the same, so both variants are handled as the same model for EC calculation. For scheduling, however, the express line buses can only be scheduled on express lines and vice versa. Note that when specific lines can be serviced by only one bus type, this practically changes an electric vehicle scheduling and charging problem with multiple vehicle types into multiple problems with a single bus type.

4.2. Selected depots

While the complete public transport region of IJssel-Vecht can be used to train the EC model, it is necessary to zoom in to depot level for the assessment of fleet schedules. As indicated by the literature review, fleet schedules can be made for a single or for multiple depots. The development of the simulation requires many iterations, so a short runtime is essential. The same holds for the generation of the alternative schedules, as four extra schedules are required per depot for the sensitivity analysis. This pleads for a simple setting in which each depot has its own, independent schedule. Multiple-depot schedules, in which buses can be exchanged between depots, are therefore beyond the scope of this case study.

Two depots are selected: Apeldoorn and Lelystad. **Apeldoorn** finds the right balance between the variation in lines and the computational effort that is required to simulate the schedule and to generate new schedules. Its schedule consists of both urban and regional lines and features multiple charger types. The network is centred around Apeldoorn Station and is considered a hub-and-spoke network, with the urban lines serving the outskirts of the city. Most regional lines have their other terminus in a nearby city and connect the towns and villages in between (figure 4.1, left).

The city lines of **Lelystad** are added as a second depot (figure 4.1, right). The choice for a second depot is made to show the effects of different charging infrastructure on the outcomes of the study. Both depots operate a hub-and-spoke network with the same bus types. However, where Apeldoorn has plenty of fast chargers, all located at the main hub, Lelystad has its slower chargers located at the depot, about six kilometres from the main hub. Every charging moment in Lelystad therefore requires a dead-heading trip, and the buses have to charge longer to recover the same amount of energy. It is to be expected that the consequences of a higher EC are felt more strongly in Lelystad than in Apeldoorn, both in the planning stage and during operations. As the results remain specific to the case, this contrast makes it possible to judge whether the conclusions hold under different charging infrastructure setups.

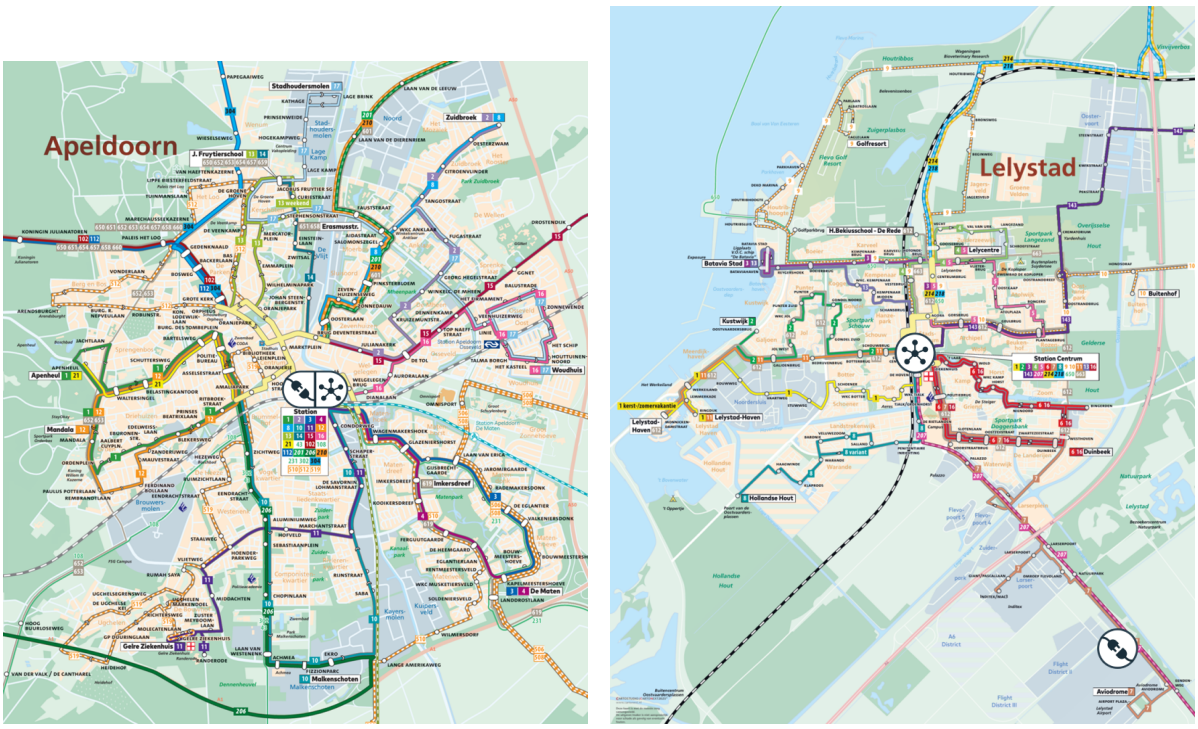


Figure 4.1: Line networks and charging locations of the depots in Apeldoorn (left) and Lelystad (right). The hub is indicated by a network symbol and the chargers by a plug symbol

4.2.1. Network, fleet and infrastructure

Table 4.1 summarises the network, fleet and charging infrastructure of both depots. The schedules cover a normal weekday (Monday to Friday, school holidays excluded). The table shows that the two depots differ considerably in size, and also in the character of their trips. The average service trip in Apeldoorn is 13.9 km long, with regional lines that are considerably longer than the city lines, whereas in Lelystad it is only 5.8 km. This matters for the EC model, as short trips are the trips for which the EC estimates are least accurate (see section 6.1.3).

Table 4.1: Overview of the network, fleet and infrastructure of the two depots in the case study

Aspect	Apeldoorn	Lelystad
Urban lines	13 (lines 1–4, 8, 10–16 and 21)	8
Regional lines	6 (lines 102, 112, 201, 206, 210, 304)	–
Service trips per day	1,137	497
Service distance per day	15,815 km	2,879 km
Service time per day	518.4 h	118.1 h
Average trip length	13.9 km	5.8 km
Vehicle tasks	45	10
Vehicle types	12m bus (35 tasks), 13m bus (10 tasks)	12m bus (10 tasks)
Battery (nominal)	12m bus: 350 kWh, 13m bus: 422 kWh, 93% state of health	12m bus: 350 kWh, 93% state of health
Main hub	Apeldoorn Station	Lelystad Station
Charging at the hub	8 $\times$ 230 kW pantograph	–
Charging at the depot	50 $\times$ 60 kW plug	10 $\times$ 100 kW plug

The charging infrastructure is the most important difference between the depots. In Apeldoorn the pantograph chargers at the station allow buses to charge during layovers at the hub, which is where most vehicle tasks pass by anyway. This is referred to as opportunity charging. In Lelystad, buses charge at the depot only, so every charging event has to be combined with a dead-heading trip. Besides charger location, the power of the chargers also differs: 230 kW at the Apeldoorn hub and 100 kW for the Lelystad depot, against 60 kW for the Apeldoorn depot. The question to what extent the outcomes are driven by the location of the chargers, and to what extent by their power, is picked up in section 6.4.9.

4.3. Benchmark and alternative schedules

The benchmark schedule is the actual schedule that was operated by EBS on normal weekdays in 2025, for each depot. New schedules are created with Hastus, a software package from the company GIRO that is specially designed for public transport planning and is used by EBS for its own schedules. The benchmark schedule was drawn up with this software, with the standard planning value for the energy consumption of each bus type. The four alternative schedules are generated with the same tool and the same settings, except for the EC that is used as input. For the schedules -20 and -10 this input is 20% and 10% lower than that of the benchmark schedule, for the schedules $+10$ and $+20$ it is 10% and 20% higher. A lower input rate means that the optimiser assumes that buses can drive further on a full battery (a longer range), which in theory results in a more efficient, but less robust schedule. A higher input rate does the opposite. The steps of 10% are chosen as they represent a balance between a considerable and a realistic change. A change of 50%, for example, would lead to interesting changes in the schedule's performance, but the chances that such a change is necessary, are low. Table 4.2 gives an overview of the schedules.

Table 4.2: Overview of the benchmark and alternative schedules. The range is the usable battery energy divided by the EC per kilometre, so a lower EC input implies a longer assumed range

Schedule	EC input (relative)	Assumed range (relative)	Vehicle tasks Apeldoorn	Vehicle tasks Lelystad
-20	-20%	about $+25\%$	45	10
-10	-10%	about $+11\%$	45	10
Benchmark	0%	0%	45	10
$+10$	$+10\%$	about -9%	45	10
$+20$	$+20\%$	about -17%	46	10

The optimisation software of EBS is treated as a black-box: its objective is to minimise an internal cost indicator, consisting of both financial costs and penalties for non-financial constraints, but its exact workings cannot be disclosed. In general terms, it solves an electric vehicle scheduling and charging problem, in which every service trip is assigned to a vehicle task (a bus), and dead-heading trips and charging events are planned in between, such that the cost of buses and kilometres is minimal and no bus runs out of energy, given a limited number of chargers. The option to invest in additional buses is kept in the optimisation, so that the effect of an adjusted input on the number of buses that are required becomes visible, even though investment decisions for the fleet are made in advance. The following

assumptions hold for all schedules: there is a flat EC rate per bus type, charging follows a linear curve, partial charging is allowed and the number of chargers is limited. The service trips are identical for all schedules, so the schedules only differ in the way the trips are linked to vehicle tasks and in the dead-heading and charging events that are scheduled in between. The alternative schedule for Apeldoorn with the highest input requires one vehicle task more (46 instead of 45). All other schedules use the same number of vehicle tasks as the benchmark.

4.4. Cost parameters

To translate the outcomes of the simulation into costs, a set of parameters is needed. These are listed in table 4.3 and are used for both depots. The values for the operator are in part derived from generalised numbers provided by the financial department of EBS, and are therefore not exact representations of the actual contract. The remainder is based on public sources.

Table 4.3: Cost parameters used in the case study

Parameter	Value	Basis
Fixed cost per bus	€125.00 per day	Capital cost of a BEB of about €45,000 per year including interest (EBS), charged for every day of the year
Variable cost per km	€0.50 per km	(EBS)
Fine for cancelled trips	€20.00 per trip	Simplified assumption, see below
Missed income	€84.00 per service hour	60% of the average cost of an hour of bus service (€140) according to Rijkswaterstaat (2026), as about 40% of EBS' revenue comes from ticket sales
Value of in-vehicle time	€7.12 per hour	Knoope (2023)
Multiplier waiting time	1.50	Wardman et al. (2016)
Multiplier late arrival	3.06	Wardman et al. (2016)
Driver cost, driving hours	€20,000 per year per daily hour	Derived from the CROW key figures (CROW-KpVV & Ecorys, 2021), adjusted to the price level of 2026, see section 6.4.6
Driver cost, idle hours	25% of the rate for driving hours	Assumption, see section 6.4.6

A few notes on these parameters are needed. The **fixed cost** is incurred for every bus that is required to run the schedule, and does not go away when the bus is not used. It is therefore multiplied by 365 days instead of by the number of operating days. The **variable cost** is the cost per kilometre that is saved when a trip is cancelled, or that is added when additional dead-heading kilometres are driven. All other cost components, such as the driver, are assumed to stay the same when a trip is cancelled, as the driver still has to be paid while sitting idle.

The **missed income** reflects the way in which most Dutch public transport concessions are paid. The operator receives a compensation (subsidy) for every hour of service that is operated, alongside the revenue from ticket sales. A cancelled trip is therefore a loss of income. The hourly amount is derived from the average cost of an hour of bus service of €140, of which about 60% is covered by subsidy. The ticket revenue is assumed to be regained, as passengers are expected to wait for the next bus.

The **fine for cancelled trips** requires more explanation. Fines differ strongly across Dutch public transport regions, and are usually calculated on the basis of the share of trips that is cancelled over a longer period. Such a percentage-based fine does not scale well with the size of a depot. A single cancelled trip in Lelystad (497 trips per day) would, for example, be penalised more than twice as strongly as in Apeldoorn (1,137 trips per day). To keep the depots comparable, a fixed amount of €20 per cancelled trip is used for both. This value is rather arbitrary, as numbers on how operators are penalised, are hard to come by. As a consequence, the fine is a modest part of the total cancellation cost (see section 6.4.5). The effect of a different fine is tested in section 6.4.7.

The **passenger cost** is based on the value of in-vehicle time (Knoope (2023)), multiplied by factors that reflect that waiting time and late arrival are experienced as more burdensome than travelling (Wardman et al. (2016)). Together, the multipliers result in a value of €32.47 per hour of passenger delay. The delay per cancelled trip is based on the headway of the line (see section 3.3.3), which is specified per line in the schedule data.

Last, **driver cost** is not part of the planned schedule cost in the main analysis, because the effect of a different vehicle schedule on the driver schedule cannot be determined without a full crew scheduling step. As personnel accounts for the largest part of the cost of an operator, the effect is illustrated in an additional analysis, using the key figures above. The first amount is the cost of an hour of driving (service and dead-heading), which differs between the schedules because of the dead-heading hours. The second amount is an assumption for the idle hours within a vehicle task (charging and waiting), which also differ between the schedules. Details are given in section 6.4.6, and the influence of the assumptions is tested in section 6.4.7.

The selection of two depots with a different charging infrastructure strengthens the case study in this respect. The differences between both depots, being fast chargers at the hub (Apeldoorn) versus slower chargers at the depot (Lelystad), give an indication of how the outcomes may change for other charging infrastructure. This is discussed further in section 6.4.9. The next chapter continues with the data that is used to train the EC model, and with the analysis of that data.

5

Data analysis

The methodology in chapter 3 describes in general terms how EC is estimated, and the case study in chapter 4 introduced the region for which this is done. This chapter forms the bridge between the two and the results. It explains how the raw data from EBS and the KNMI is turned into a data set that is suitable for the regression model, and analyses that data set before any model is fitted. The analysis answers three questions. How good is the quality of the data, and where are its limits? How are the values of every factor distributed, and what does that mean for the model? And how do the factors relate to each other and to EC? The answers shape the choices that are made in the modelling and help to interpret the results in chapter 6.

5.1. Data processing

As stated in the methodology, data is provided by two sources. The weather data is openly available through the website of the KNMI. The weather station that lies closest to the centre of the IJssel-Vecht area is Heino, Overijssel. The data is provided in hourly intervals, so there are 8,760 instances over the whole of 2025. The wind direction is given in degrees, but when there is little to no wind, the value is set at 990. As these values do not represent a valid measurement, they are corrected through linear interpolation between the neighbouring hours. This affected 158 of the 8,760 records. Every trip is assigned weather data based on the hour in which the trip started.

The data provided by EBS consists of two parts, both recorded in electric buses in the IJssel-Vecht transport region during 2025. The first part holds for every trip the origin and destination stop, the departure and arrival time and the vehicle number. The second part holds the consumption data of that same trip: the bus type, the scheduled distance, the highest and lowest SoC recorded during the trip, the battery capacity, the SoH and the number of passengers (*pax*). Due to flaws in the data collection, not all trips during the year are included, but the amount of data can be considered great nonetheless. The initial data set for EC and bus specifications contained 1,048,753 trips and the set with information about travel times had 1,164,736 unique records.

To arrive at the factors that are used in the regression model, several variables are derived from this raw data:

- **EC** is provided as the difference between the highest and lowest SoC value during a trip. The amount of kWh follows from multiplying that difference with the battery capacity and the SoH. As explained in the methodology, the EC is expressed per kilometre, by dividing it by the scheduled distance of the trip.
- **Max SoC** is in this study equal to the initial SoC at the start of a trip.
- **Average speed** is calculated from the actual travel time and the scheduled distance. The **difference in travel time** is the actual travel time minus the scheduled travel time, so that negative values indicate that the bus was faster than planned.

- **Route type** is assigned on the basis of the line overview. Lines within a city are marked as city lines (0), all other lines as regional lines (1).
- **Headwind** is the component of the wind speed in the direction of travel. The direction of a trip is approximated by the bearing between the coordinates of the origin and the destination stop. The headwind then follows from multiplying the wind speed with the cosine of the difference between wind direction and bearing. A positive value indicates headwind and a negative value tailwind. The wind direction is the interpolated direction described above.
- **Pax** is derived from check-in data. This holds information about how many unique passengers boarded the bus on that trip. The advantage is that the number is available for a large number of trips. The disadvantage is that it does not say anything about the occupancy over the course of the trip, as it is unknown how long each passenger rode along. This is unfortunate, as the average occupation rate, and with it the extra weight on board, is what would be preferred for the calculation of EC. For approximately 10% of the trips (97,315) the number of pax was not available. A more detailed analysis showed that most of these instances were either the first or the last trip of the day on a line, or trips during holidays. There tends to be less demand for these trips in general, so the assumption is made that these trips did have zero pax.

The data is cleaned in several steps. Trips with incomplete EC records are removed, as are trips for which no travel time information could be linked. Next, trips that are physically implausible or that are considered outliers are filtered out: trips with a travel time of zero or a deviation from the schedule of more than 20 minutes, an average speed outside the range of 10 to 60 km/h, an EC of 2.5 kWh/km or higher, more than 10 pax per kilometre of scheduled distance, or no recorded energy use on a trip longer than 10 km. Table 5.1 lists the number of trips that remain after every step. In total, 884,860 trips remain, which is about 84% of the original EC records. Most of the trips were lost because no matching travel time record was available. The EC model is trained for the two bus types that feature in the case study. This leaves 666,761 trips for the 12m bus and 50,350 trips for the 13m bus. The other trips belong to bus types that are not used by the depots of the case study.

Table 5.1: Number of trips remaining after every step in the data processing

Step	Trips removed	Trips remaining
Raw EC records (2025)	–	1,048,753
Incomplete EC records	5,558	1,043,195
Missing travel time data after linking	151,083	892,112
Implausible trips (travel time, speed, EC and pax)	7,252	884,860
12m bus (used for the EC model)		666,761
13m bus (used for the EC model)		50,350

The result is a large data set with, at first sight, good coverage of all seasons and lines. The next section looks at the quality of the EC data, because the way in which EC is measured has consequences for the rest of the analysis.

5.2. Data quality

The flaws that remain in the data are expected to be countered by the sheer amount of valid data points. However, one property of the data requires specific attention. The SoC is recorded as an integer percentage of the battery capacity. The EC of a trip can therefore only increase in discrete steps. With a usable capacity of about 325 kWh for the 12m bus and 392 kWh for the 13m bus, one step is equal to roughly 3.2 and 3.9 kWh respectively. For a long trip of 40 km, a single step amounts to less than 0.1 kWh/km and the measurement is reasonably precise. For a short trip of 5 km it is a different story: the EC per kilometre can then only take values that are multiples of 0.65 kWh/km.

This is clearly visible in figure 5.1. The measured EC per kilometre follows distinct curves, that each correspond with a fixed number of energy steps divided by the distance of the trip. The curves lie further apart for shorter trips, which means that the measurement error on short trips is large. This has implications for the 12m bus in particular, as a large share of its trips is short (63.4% of the trips is shorter than 10 km). For the 13m bus, which mostly operates on regional lines, the effect is less pronounced.

The consequence is that the EC model will be less accurate for individual short trips, regardless of how well the factors are chosen. This is considered in the validation, where the performance of the model is also assessed over a complete vehicle task (see section 6.1.4).

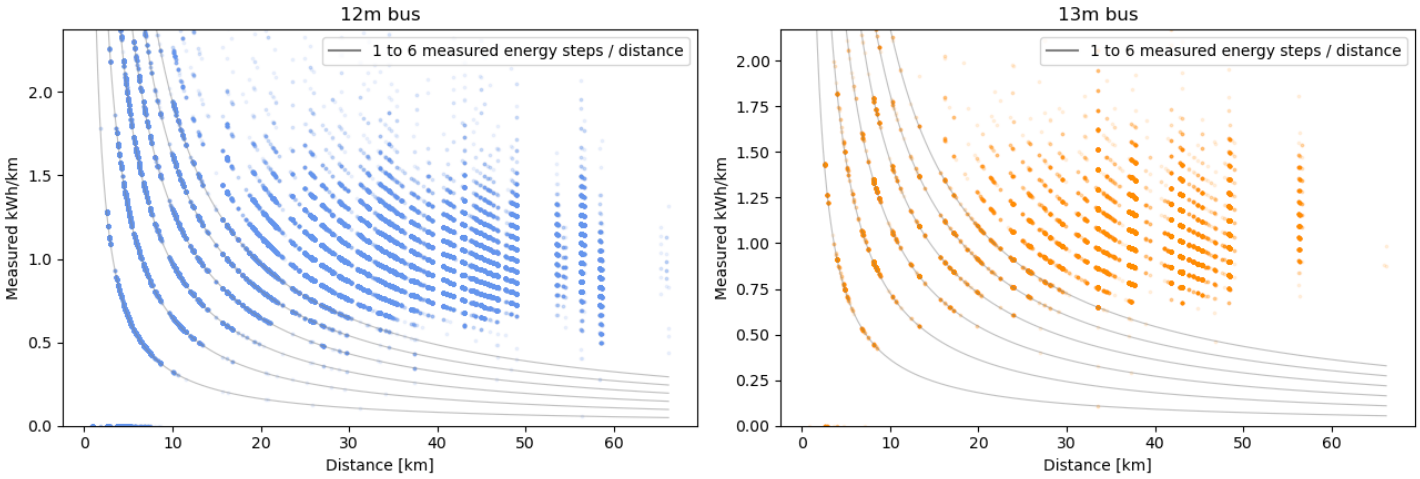


Figure 5.1: Measured EC per kilometre plotted against trip distance. The grey lines indicate the EC that corresponds to one up to six discrete energy steps

With the data cleaned and its limitations known, the next step is to look at what the data actually contains. The following section describes the distribution of the values within each factor.

5.3. Factor value distributions

Before the relationship between the factors and EC is studied, it is important to know in which range the values of each factor occur. Figure 5.2 shows the distribution of every factor for both bus types. Binary factors are shown as a share of trips, continuous factors as a density. The displayed range of the continuous factors is cut off at the 0.01^{th} and 99.99^{th} percentile to keep the figure readable. The observations outside that range (at most 0.07% per factor) are still part of the data set.

Route type and scheduled distance. Route type coincides almost entirely with bus type for the 13m bus: nearly all of its trips are made on regional lines. The small set of short city trips is caused by buses that occasionally drive other routes than planned. The 12m bus is used on both types of lines, but mostly on city lines. This is also visible in the distribution of the scheduled distance. The 12m bus has a high share of short trips, with a peak between 5 and 7 km, whereas the 13m bus hardly makes short trips and has several peaks between 30 and 50 km that correspond with specific regional lines.

Difference in travel time and average speed. The difference in travel time follows a normal distribution, in which the largest part of the trips is faster than scheduled. The average speed is the speed that was actually driven. For the 12m bus, the highest concentration lies between 20 and 30 km/h, which fits the many stops on city lines. The 13m bus shows a clear peak around 35 km/h. The small peak around 60 km/h is likely caused by trips over motorways. As trips with an average speed of more than 60 km/h were filtered out, this peak is cut off.

Pax. The number of pax follows a distribution that resembles an exponential or log-normal distribution. Most trips have only a few passengers, while a small number of trips is very busy. This also implies that the extremely high values are based on a limited number of observations.

Max SoC. The max SoC shows a remarkable pattern, with two outliers at which the density jumps. It is not clear what causes these. The pattern itself can be explained, though. When every bus starts the day fully charged, there is always a trip that starts around 100%, while lower values are reached less often. The initial SoC also rarely seems to drop below 20%. This is probably the result of the way in which buses are scheduled, with a margin to prevent running out of energy.

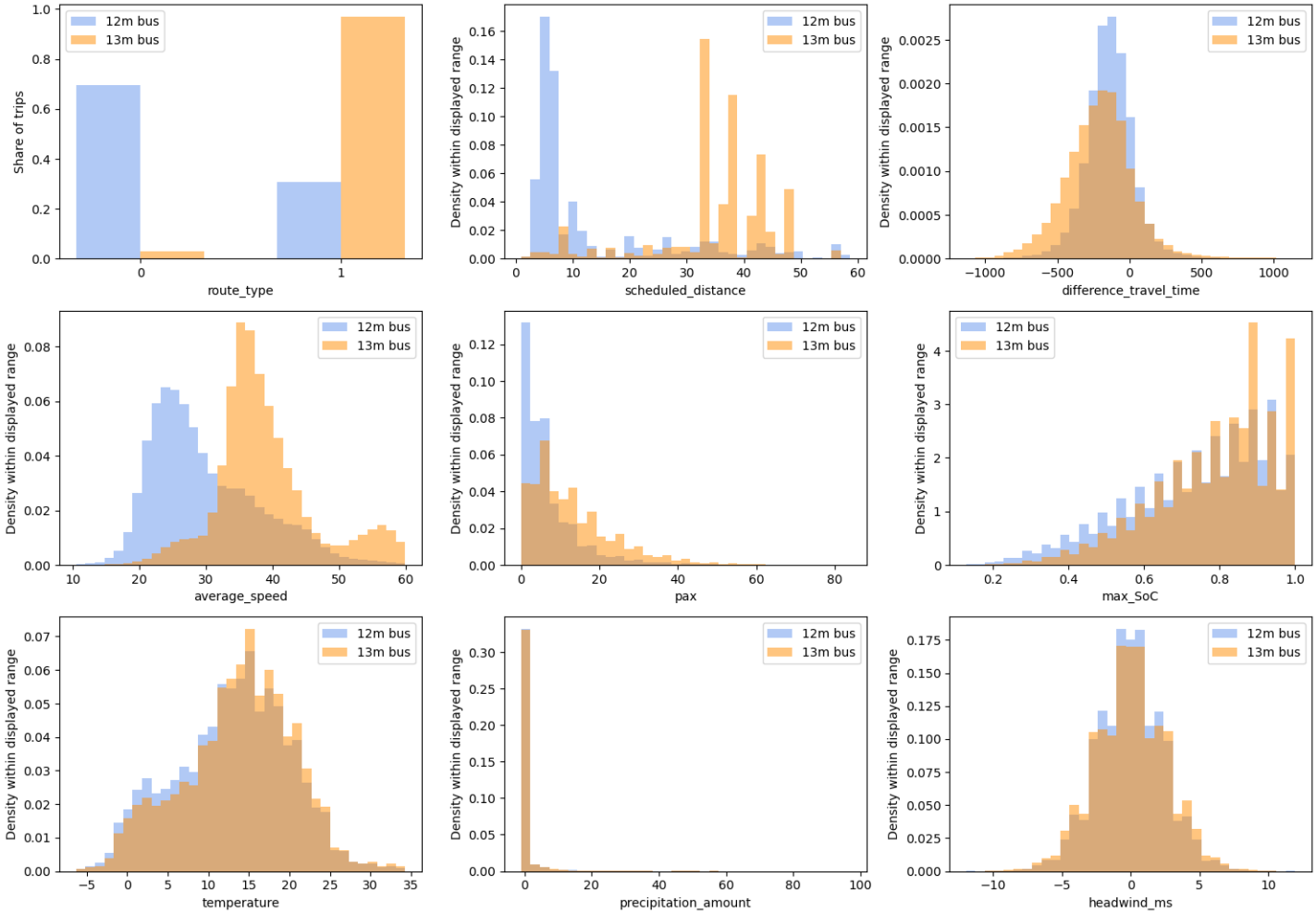


Figure 5.2: Distribution of values within every factor for both bus types

Temperature, precipitation and headwind. Most trips are made at a temperature around 15°C. The left tail of the distribution is much thicker than the right tail, so low temperatures occur more often than high temperatures, which is relevant for the choice of scenarios later. Precipitation shows a strong concentration around 0 and 1 mm/h. As it is a translation of the number of millimetres that fall, there is a good chance that the factor has little influence on the model. Headwind follows a normal distribution, with a strong fluctuation in it. In most cases there is only a light headwind or tailwind.

An important consequence of these distributions is that the factors cannot be treated as independent. The values in the distributions are related, as will be shown in the next section, and the distributions are also the result of the way in which the schedules are built. It is therefore too early to conclude that a random value can be drawn from each distribution to analyse future robustness. This is further elaborated in section 6.2.

5.4. Factor correlations

The relationship between the factors themselves is relevant to measure, because correlated factors can eat away each other's effect in a regression model. If two factors describe a similar effect, the model can attribute the effect to either one of them. This makes the individual coefficients, and with them the interpretation of the importance of a factor, less reliable. To assess this, the Pearson correlation coefficient is calculated for every combination of two factors. The coefficient ranges from -1 to $+1$, in which a value of 0 means that there is no linear relationship between the factors. A coefficient of 0.7 or higher (in absolute terms) is generally considered as strong, and is used as the threshold here. The

results are presented in figure 5.3.

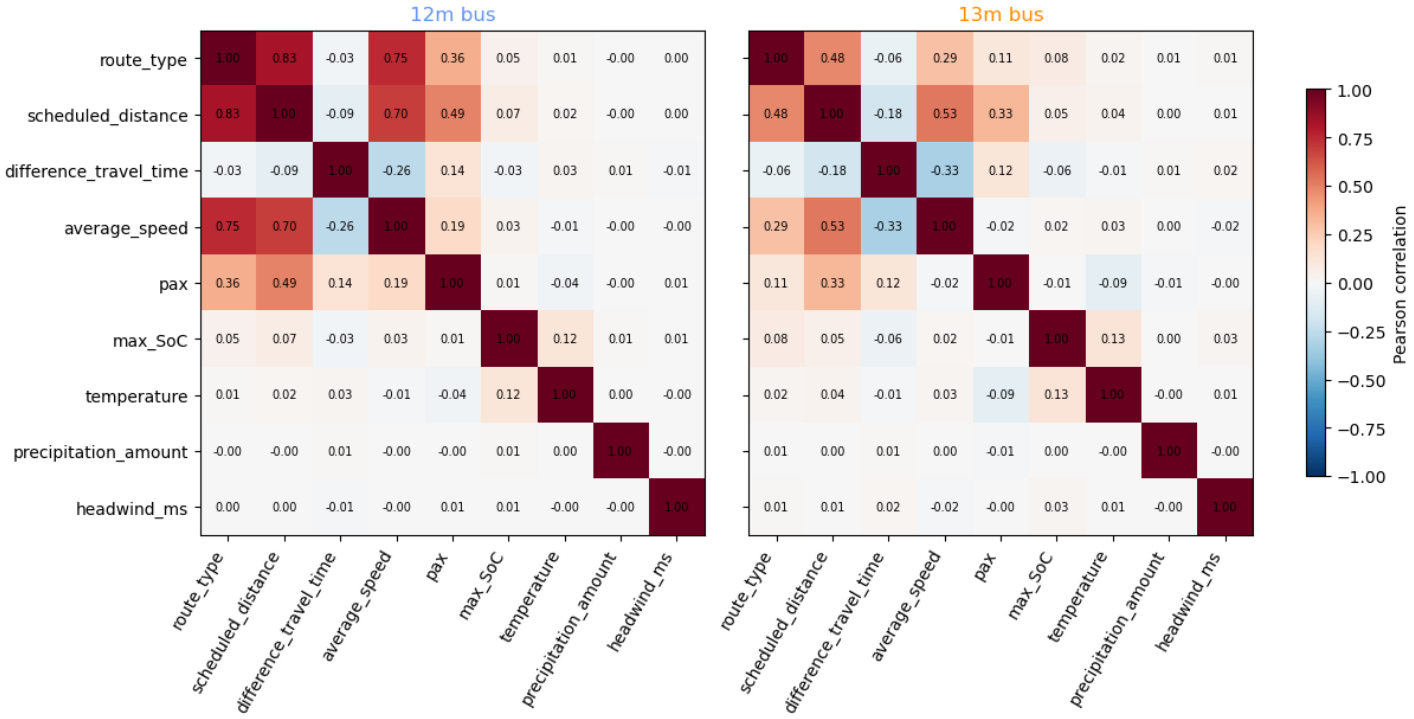


Figure 5.3: Pearson correlation coefficient for each combination of two factors

For the **12m bus** route type correlates strongly with scheduled distance (0.83) and average speed (0.75). At the same time, scheduled distance and average speed also correlate strongly with each other (0.70). The reason is that the difference in route type implies other characteristics as well. A city line often covers a smaller distance, as it connects neighbourhoods within the same city. In addition, a bus can reach a higher average speed on regional roads than on city streets. This is partly due to the speed limits, but probably also to the amount of intersections and stops at which the bus has to wait. Pax shows a moderate correlation with route type and scheduled distance (up to 0.49). This is logical, as more passengers can board and alight on longer lines. The remaining factors show little correlation. This benefits the usability of the estimations for those factors. For the factors that were mentioned above, it is a matter of taking these insights into account when the effects are assessed in the results.

For the **13m bus** the pattern is similar, but the coefficients are lower. This is in line with the observation that this bus type is almost exclusively used on regional lines, where there is less variation in route type and the other two factors. The correlations are not strong enough to remove a factor from the analysis. They do mean that the effect of route type, scheduled distance and average speed has to be interpreted as a group. In section 6.1.2 this comes back when the importance of the factors is discussed.

5.5. Conclusions from the data analysis

The data analysis leads to three conclusions that are taken to the results. First, the quality of the data is sufficient for a regression model, as the sheer number of valid trips counters the remaining flaws. Second, the effects of route type, scheduled distance and average speed are closely connected, so they have to be interpreted as a group when the outcomes of the model are discussed. Third, the integer SoC makes the EC of short trips unreliable. It is therefore useful to validate the model not only on the level of trips, but also over multiple trips. The next chapter uses these insights to fit the EC model, to determine the relationships and the importance of the factors, and to validate the performance of the model.

6

Results of the case study

The data analysis has shown how the raw data is turned into a data set for the EC model, and which patterns it contains. In this chapter the methodology is applied to the case study to answer the sub-research questions posed earlier. The outcomes are tied to the case study, but an effort is made to generalise the findings. Each sub-question has its dedicated section, and each section builds upon the knowledge gained in the previous parts. Instead of a separate discussion, the outcomes are compared with the literature, and the limitations are discussed, in the section to which they apply. Every section ends with a short conclusion, which answers the corresponding sub-question. The chapter starts with the EC model, which answers the first sub-question (section 6.1). Based on the most influential factors, the scenarios are defined in section 6.2, which answers the second sub-question. Section 6.3 uses these scenarios to assess the benchmark schedules, which answers the third sub-question. Finally, section 6.4 compares the benchmark schedules with the alternative schedules, to answer the fourth sub-question.

6.1. EC model

The EC model is at the heart of this research. The first step is to understand the relationships between every factor and EC. Then the importance of each factor is determined. Because the model is used to explain the consequences of choices made in the planning phase, an accurate model is preferred. The last part of this section is therefore dedicated to evaluating the model performance, both on trip level and on vehicle task level. The model is a polynomial regression of degree two for the EC per kilometre, trained for the 12m bus and the 13m bus separately with five-fold cross-validation. All metrics presented below are calculated on the out-of-fold predictions, so that every trip is predicted by a model that has not seen that trip.

6.1.1. Factor relationships and importance

The fitted model for each bus type is a degree two regression on the nine factors, so with the factors x_{ij} of trip i,

$$\widehat{y}_i = \beta_0 + \sum_{j=1}^{9} \beta_j x_{ij} + \sum_{j \leq k} \beta_{jk}\, x_{ij} x_{ik}, \tag{6.1}$$

which results in 54 terms (nine linear, nine squared and 36 interaction terms) and an intercept per bus type. Table 6.1 lists the estimated coefficients of the linear and squared terms, in the order of the importance of the factors that is derived later in this section. The factors are standardised, so a coefficient is the effect in kWh/km of a change of one standard deviation, at the mean value of the other factors, and the coefficients can be compared between factors. The intercept is the EC of a trip for which all factors have their mean value. The coefficients of the 36 interaction terms are not shown. The squared coefficient of temperature is positive, in line with the U-shaped relationship, and together with the large negative linear coefficient it makes temperature the dominant term of the model. The

linear coefficients are conditional on the squared terms and the correlations between the factors, so they should not be read as the effect of a factor on its own. The effect of a factor over its range is better read from the partial dependence plots, which are discussed next.

Table 6.1: Estimated coefficients of the linear (β_j) and squared (β_{jj}) terms of the EC model, in kWh/km per standard deviation of the factor. The intercept β_0 is 0.8596 for the 12m bus and 0.9140 for the 13m bus. The 36 interaction terms are not shown

	12m bus		13m bus	
Factor	β_j	β_{jj}	β_j	β_{jj}
Temperature	-0.1056	$+0.0624$	-0.0755	$+0.0481$
Max SoC	-0.0078	$+0.0345$	$+0.0016$	$+0.0263$
Pax	$+0.0673$	-0.0039	$+0.0486$	-0.0039
Average speed	-0.0056	$+0.0074$	$+0.0023$	$+0.0044$
Difference in travel time	$+0.0307$	$+0.0007$	-0.0039	$+0.0038$
Headwind	$+0.0161$	$+0.0104$	$+0.0241$	$+0.0083$
Route type	-0.0128	-0.0108	-0.0013	$+0.0073$
Scheduled distance	-0.0284	$+0.0142$	-0.0128	$+0.0210$
Precipitation	$+0.0175$	-0.0009	$+0.0163$	-0.0008

The degree of the regression model is 2, implying that both linear and quadratic relationships are possible. Figure 6.1 shows the relationship for every factor in a partial dependence plot. This analysis method sets all other factors to their mean value and only changes the input of the relevant factor. In that way, it becomes visible how the EC reacts to a change in the factor value. Each plot also marks the distribution of values within the factor by dividing the data in 10 bins. The border between two bins is depicted as a small vertical line. When the lines are close to each other, the values are concentrated around these levels and vice versa. A high concentration leads to smaller differences between the values and therefore the factor is also less likely to influence the EC. Extreme values are indicated by the 5^{th} and 95^{th} percentile. The plots are complemented by tables 6.2 and 6.3, that show for each bus type the lowest and highest EC within the 5^{th} to 95^{th} percentile range of every factor, the factor values at which these occur and the resulting difference (also expressed as a percentage of the lowest EC). The factors are ranked on that difference.

Table 6.2: 12m bus – PDP range within 5th–95th percentile range

	Lowest EC		Highest EC		Difference EC	
Feature	**EC value**	**factor value**	**EC value**	**factor value**	**EC value**	**percentage**
temperature	0.868	18.98	1.285	0.70	0.417	47.98%
pax	0.918	0.00	1.096	22.00	0.178	19.40%
max_SoC	0.939	0.76	1.079	0.40	0.140	14.90%
difference_travel_time	0.924	-401.00	1.016	121.00	0.091	9.88%
headwind_ms	0.954	-1.94	1.015	3.96	0.061	6.38%
route_type	0.927	1.00	0.973	0.00	0.046	4.94%
scheduled_distance	0.956	28.39	0.999	3.97	0.043	4.46%
average_speed	0.954	31.61	0.976	45.38	0.021	2.22%
precipitation_amount	0.969	0.00	0.980	3.00	0.012	1.20%

Table 6.3: 13m bus – PDP range within 5th–95th percentile range

	Lowest EC		Highest EC		Difference EC	
Feature	**EC value**	**factor value**	**EC value**	**factor value**	**EC value**	**percentage**
temperature	0.932	19.32	1.261	1.20	0.328	35.22%
scheduled_distance	1.005	38.10	1.184	8.48	0.180	17.90%
pax	0.963	1.00	1.100	34.00	0.138	14.32%
max_SoC	0.986	0.79	1.079	0.49	0.093	9.38%
headwind_ms	0.984	-3.83	1.060	4.05	0.075	7.66%
average_speed	1.005	36.36	1.034	55.76	0.028	2.82%
difference_travel_time	1.007	-72.46	1.024	-600.00	0.017	1.71%
precipitation_amount	1.009	0.00	1.023	4.00	0.015	1.44%
route_type	1.012	1.00	1.012	1.00	0.000	0.00%

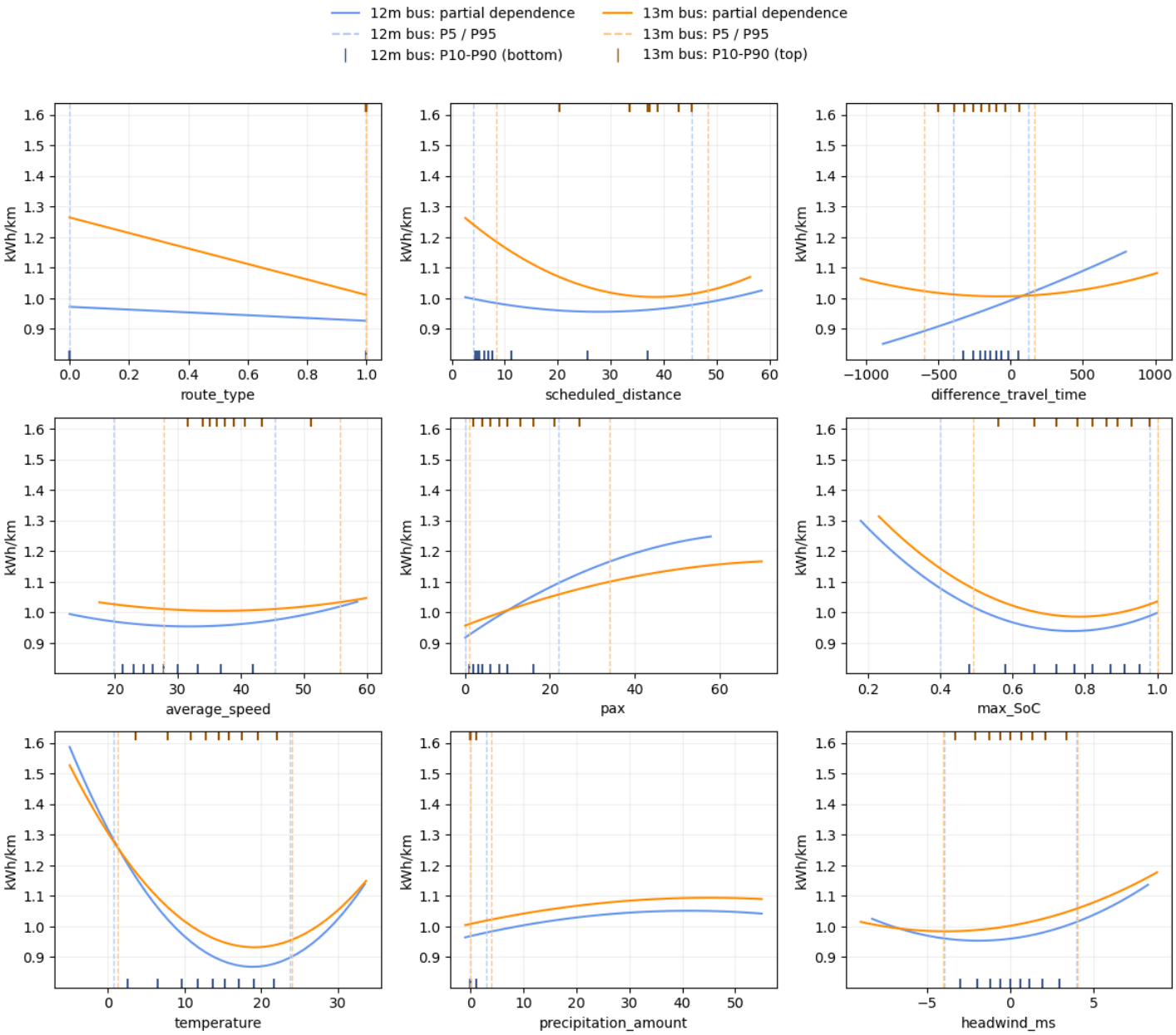


Figure 6.1: Partial dependence plots for all nine factors included in the model

The relationships are discussed factor by factor. Route type, scheduled distance and average speed are treated first, as their strong correlations (see section 5.4) come back in each individual explanation.

Route type shows a small difference between city lines (0) and regional lines (1) for the 12m bus, with the regional lines having a slightly lower EC. For the 13m bus the difference is greater, but the number of trips on city lines is too low to draw meaningful conclusions. As the factor only has two options, the relationship is linear by definition. The difference is probably caused by the fact that multiple characteristics change with the route type, which was also the main reason for including it in the model.

Scheduled distance should theoretically not impact EC when the EC is measured per unit of distance. The high correlation with average speed and route type shown in section 5.4 suggests that the factor takes over a part of their effect. It could also be that scheduled distance correlates with another factor that is not included in this research. Because the correlation coefficients are more modest for the 13m bus, the factor covers a larger range of EC values for that type, although measured trips on those

distances occur less frequently.

Average speed is the most theoretically sound factor of the three. For the 12m bus it has a quadratic relationship with a minimum at 33.30 km/h. This can be seen as a balance between low speeds, at which frequent stopping causes the auxiliary systems, such as the HVAC, to consume more power over the same distance, and high speeds, at which the drivetrain consumes more energy. For the 13m bus the relationship between average speed and EC is almost flat. The **difference in travel time** has a linear effect on EC for the 12m bus. Trips that take less time than planned record a lower EC, probably for the same reason as with average speed. For the 13m bus this is not the case. Here higher speeds could mean that the bus consumes more energy as drag increases more rapidly at higher speeds.

The remaining partial dependence plots show similar patterns for both bus types within the 5^{th} and 95^{th} percentile range, so the explanation of the two models is combined. In most cases the 13m bus registers a slightly higher EC rate.

Pax follows a largely linear pattern within 90% of the values. The slight bend towards higher values might be due to outliers in the data or to the fact that pax is unevenly distributed in trips with high pax. The relationship can indicate extra power required due to the weight added by passengers, but it can also tell something about the number of stops that the bus had to make for boarding and alighting. If a bus has more unique passengers on board, the chance that the bus has to stop at more bus stops grows. As pax is based on check-ins, the two explanations cannot be separated.

Max SoC has a quadratic pattern with a minimum at 77%. The EC increases at both ends, but in a more extreme fashion at the left side of the graph. At the same time, there are fewer observations with a low SoC, so the predictions there are somewhat uncertain. A plausible explanation for the higher EC at near full SoC levels, is that the buses in this study cannot be heated or cooled during charging. The bus therefore has to do this immediately after it stood idle for a while. The first trip is impacted more by the additional energy required to power the HVAC systems. The steep increase at the lower end of SoC values can be caused by so-called SoC drops. This is a phenomenon where the battery registers a higher remaining capacity than what is actually the case. When the bus reaches a lower SoC, the battery management system has to adjust the SoC level, which results in a higher recorded consumption. This is a problem in both the way the EC data is gathered, as well as the ability of the battery management systems to fully grasp the battery's current state.

Temperature shows the largest difference in EC rates. The relationship is again quadratic, with a minimum at 19.01°C. In the Dutch climate (at least in 2025) low temperatures are recorded more often than high temperatures. This makes the prediction of the model more certain on the left side of the parabola. The explanation for this, as frequently researched in other studies, has to do with both the chemical nature of the batteries, that lose capacity as the temperature deviates further from the optimum at room temperature, and with the increased power consumption of the HVAC. The aim of the heating and cooling equipment in the bus is to keep the temperature around 21°C. When temperatures decrease, the heating has to work harder and needs to consume more energy, and vice versa for cooling at high temperatures.

Precipitation is, as was already mentioned in the data analysis chapter, very concentrated around values below 5 mm/h. The relationship shows that there is little influence on EC between those low levels. This could be explained by the fact that falling rain has little direct effect. When rain falls in consecutive hours and causes standing water on roads, it might have an effect on EC, but this is not taken into account at the moment.

Headwind has a quadratic relationship, where EC increases not only when headwind increases, but also when headwind decreases from $-3.25\ \mathrm{m}_{/\mathrm{s}}$ onward. This unexpected behaviour is hard to explain exactly. The most plausible explanation is that the bus suffers from higher drag the more wind there is, no matter the direction. Furthermore, the headwind is a simplification that is based on the general direction between the start and end node, which is unable to capture variations in direction throughout the trip. The wind measurements are also taken from a weather station that is located somewhere in the centre of the case study area. There might be relatively large differences between lines within the case study. The same can be said about precipitation.

The stark increase in EC at temperatures that occur relatively frequently in the Netherlands makes tem-

perature a prime candidate for the scenario analysis. To confirm this position, and to find other factors that should be included, the next section discusses the importance of the factors in this regression model.

6.1.2. Factor importance

In addition to the relationships, the importance of each factor is analysed. This is done by assessing the relative difference in EC rates in the partial dependence plots, and by doing two related tests on the R^2. The first test randomly shuffles all the values in a factor (permutation) and the model is run again. The difference in the R^2 tells how the model performed before and after the shuffle. The impact can be larger than the original model R^2 as the shuffle might lead to a model that is worse at predicting the EC than random guessing. The second test is to leave out the factor completely (drop-column) and to refit the model. In this way, the R^2 cannot drop below 0 and thus the differences are smaller.

Both tests matter, as they answer a different question, and the combination says something about the way in which the model uses a factor. The PDP tables show how much EC changes over the range of a factor, but not how much of the variance in EC is explained by it. The shuffle test shows how much the model relies on a factor as it is currently specified. The drop-column test shows how much information is lost when the factor is not available at all. For a factor that is largely redundant, because another factor contains similar information, the drop in R^2 is small when it is left out, while the drop is large when its values are shuffled. In that way the tests also help to interpret the correlations found in the data analysis. As the drop-column value is the most robust indicator for the added value of a factor, it determines the ranking of the factors in figure 6.2. The list is sorted on the impact of dropping a factor, weighted by the number of trips of both bus types.

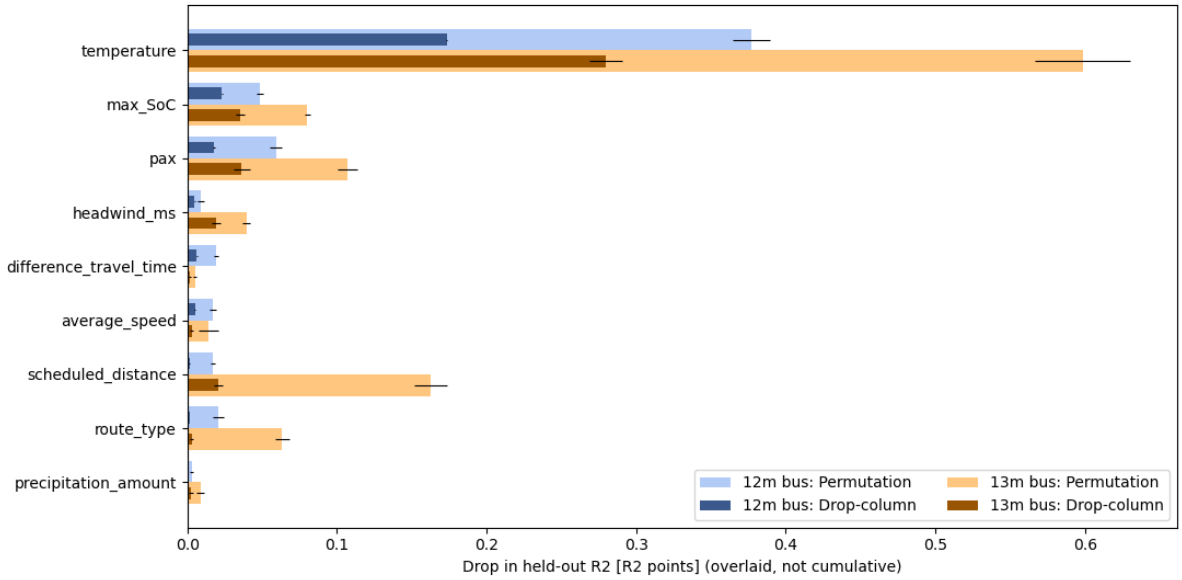


Figure 6.2: Drop in held-out R^2 when the values of a factor are shuffled (light bars) or when the factor is left out of the model (dark bars), for the 12m bus (blue) and 13m bus (orange)

Figure 6.2 shows that the ranking of the factors is fairly similar for both bus types. The one difference that must be noted is the scheduled distance, which is discussed below.

Temperature is in every test the most important factor by a margin. Leaving it out reduces the R^2 by 0.173 for the 12m bus and by 0.279 for the 13m bus, which is about a factor seven to eight more than the next factor. This confirms the conclusion from the partial dependence plots, and aligns with the literature on the EC of electric buses. **Max SoC and pax** then follow, with a drop of 0.023 and 0.018 for the 12m bus, and 0.035 and 0.036 for the 13m bus. This indicates their relevance for the simulation. In the case of the max SoC this can be incorporated by taking the SoC at the end of the previous trip as the max SoC of the trip thereafter. For pax this is less obvious. How this factor is included in the

scenarios is discussed in section 6.2.

Difference in travel time, average speed and headwind lead to about the same drop in R^2 (0.005 to 0.006 for the 12m bus) when each of them is excluded from the model. The light bar shows a smaller drop for headwind than for difference in travel time and average speed when the values are shuffled, which means that the model relies less on headwind as it is specified. For the 13m bus, headwind is more important (0.019), which is in line with the higher PDP range for that bus type in table 6.3. A possible explanation is that the 13m bus is mostly used on regional roads with higher speeds, where the effect of wind is larger.

Route type and scheduled distance show to have little effect on the overall model performance when they are left out of the 12m bus model (0.002 and 0.001), but the effects are far greater when their values are shuffled (0.021 and 0.017). This indicates that both factors could have a larger effect on their own, but that in the complete model the effect is taken over by average speed and the other factors with which they are correlated. This is in line with the strong correlations found in the data analysis, and gives insight into how the effects of correlated factors are distributed over the model. As the tests do not allow to separate the effect of the individual factors in this group, the conclusion should be seen as an interpretation, rather than as proof.

The scheduled distance does behave differently for the 13m bus. Here, both the drop-column (0.021) and the shuffle value (0.163) are high, so the factor contains information that is not available elsewhere in the model. There are several possible explanations. The correlation with route type and average speed is lower for this bus type, so there is less overlap. Furthermore, the trips of the 13m bus are distributed over a limited number of regional lines with a specific length (see the distribution in section 5.3). Scheduled distance can therefore act as an indicator of the line, and capture line-specific characteristics, such as the number of stops, the type of road or the height difference, that are not included as separate factors.

Once again, **precipitation** ranks below all other factors and has almost no impact on the model (a drop of 0.001 to 0.002). Precipitation will therefore not be included in the scenarios.

Besides the individual relationships between the factors and EC, a second degree model also includes interaction effects between two factors. Whether these effects change the picture is discussed next.

Interaction effects

An interaction effect means that the effect of one factor on EC depends on the value of another factor. In a linear model without interaction the effect of a factor would be equal for all values of the other factors. Figure 6.3 illustrates this for the combination of average speed and temperature, which is the strongest interaction for the 12m bus. At low temperatures, EC decreases with increasing average speed. At high temperatures the effect is the other way around: EC increases with speed. A possible explanation is that the energy for heating is spread over fewer kilometres when the bus drives faster, so a higher speed is beneficial in cold weather. In warm weather, the demand for heating or cooling is smaller and the drag, which increases with speed, becomes dominant. For the 13m bus the curves for low, middle and high temperature are more parallel, and the interaction is therefore smaller.

Table 6.4: Top five interaction effects with the largest gain in held-out R^2 (compared to a model with only main and squared terms)

Bus type	Interaction	Mean fold R^2 with interaction	Gain
12m bus	average speed × temperature	0.2498	0.0040
	route type × pax	0.2492	0.0033
	scheduled distance × pax	0.2485	0.0027
	scheduled distance × difference in travel time	0.2484	0.0026
	route type × difference in travel time	0.2483	0.0024
13m bus	scheduled distance × pax	0.4553	0.0052
	average speed × pax	0.4518	0.0017
	route type × scheduled distance	0.4512	0.0012
	pax × max SoC	0.4510	0.0010
	route type × difference in travel time	0.4510	0.0010

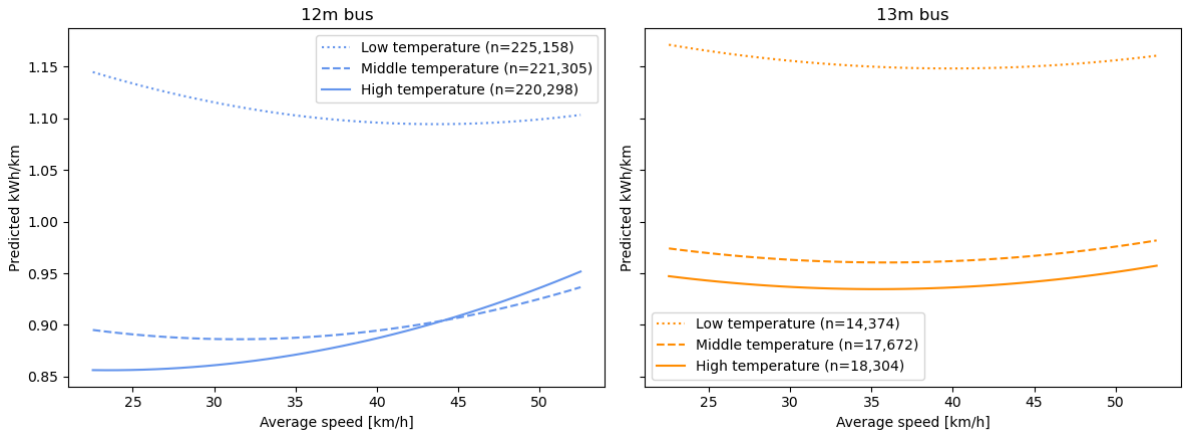


Figure 6.3: Example of the interaction effect between average speed and temperature, for the 12m bus (left) and 13m bus (right)

Because the effect of the interactions is small, the individual relationships between the factors and EC are a good representation of the model. This is favourable for the interpretation of the relationships, and it makes it more likely that the difference between the drop-column and the shuffle test can be explained by correlated factors that capture each others effect, rather than by the interactions. With the relationships and the importance of the factors established, the remaining question is how well the model predicts EC.

6.1.3. Validation on trip level

The regression model estimates the EC on trip level. To validate how well the model performs, the residuals (in kWh/km) of the trips are assessed. The residual is the difference between the predicted and the actual EC, so a positive residual means that the model overestimates the EC. The relative difference divides the difference by the actual value, to get the deviation as a percentage. This is useful to assess the residuals across all distances in the same way. All residuals are out-of-fold, meaning that they are predicted by the model that was not trained on that trip.

Figure 6.4 shows the performance through both a scatter plot and a histogram of the residuals. Especially the 12m bus shows a wide range, which can be problematic. A robustness analysis that, in itself, has a high uncertainty, can hardly be used to draw conclusions. The 13m bus shows a much more workable pattern, although a deviation of 0.2 kWh/km from a mean of 1.0 kWh/km still indicates a relatively high uncertainty. In addition to the visual representation, table 6.5 shows the metrics that indicate the model's accuracy. The R^2 is 0.27 for the 12m bus and 0.47 for the 13m bus. The 5^{th} and 95^{th} percentile of the residuals are -0.497 and $+0.389$ kWh/km for the 12m bus, which is a wide margin compared to the average EC of about 1 kWh/km.

The main difference between the trips of the 12m buses on one hand, and the 13m buses on the other hand, is the route type, average speed and scheduled distance (see figure 5.2). The high correlation between the three factors shown in the data analysis explains why these are named together. These insights, in combination with the fact that EC data is measured in discrete steps of roughly 3.2 to 3.9 kWh, lead to the hypothesis that short trips suffer heavily from rounding errors. Figure 6.5 supports this hypothesis. Trips shorter than 10 km make up 63.4% of the 12m bus trips and only 6.0% of the 13m bus trips, and they account for the largest part of the wide tails of the residual distribution of the 12m bus.

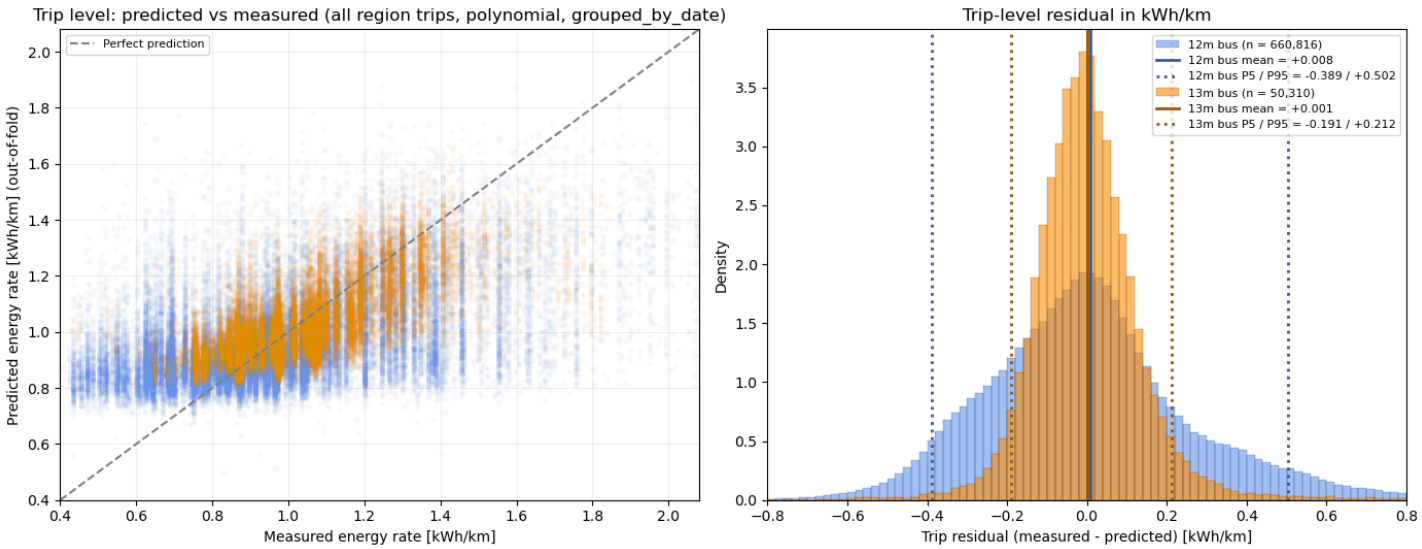


Figure 6.4: Scatter plot (left) and histogram (right) with the trip level residuals

Table 6.5: Model accuracy metrics on trip level

Vehicle type	R^2	RMSE (kWh/km)	MAE (kWh/km)	P5 residual (kWh/km)	P95 residual (kWh/km)
12m bus	0.2660	0.2688	0.2011	-0.4970	0.3885
13m bus	0.4691	0.1357	0.0951	-0.2024	0.1860

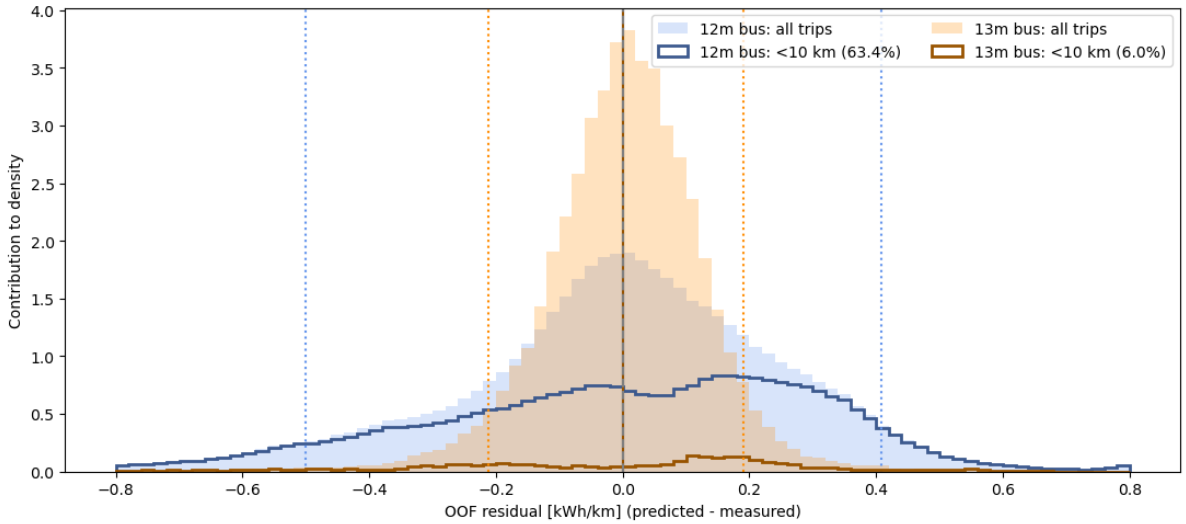


Figure 6.5: Effect of trips with a distance below 10 kilometres on the residuals. The dotted lines indicate the 5^{th} and 95^{th} percentile of the residuals of all trips

The plots where trip residuals are set off against the factor values (figure 6.6) are designed to analyse whether the model has the ability to capture non-biased effects with the currently estimated relationships. Systematic deviations are indicated by the (sometimes hard to see) orange and blue lines that hover around the residual line. The line is based on the mean value in a certain percentile bin. There are 10 bins in total. An unbiased model should lead to predictions where the residuals in each bin are equally distributed around the prediction, and thus the blue and orange line should be drawn at a residual of 0.

There are three relatively large deviations that can be spotted in the graphs. First, the bin mean fluctuates quite strongly for bins with a short distance, indicating that the model has difficulties finding patterns in short trips. Second, at high max SoC values both lines over- and undershoot the zero line. This means that the linear or quadratic relationship does not fit well at those values. For the 12m bus it also shows a dip at medium levels, which points to a systematic overestimation there. In other words, the model estimates a higher EC than the actual value. Third, a similar overestimation can be found at temperatures of around 3°C for both models. These biases should be taken into account when looking at the relationships earlier in this chapter.

The plots largely show that the model does not feature a significant bias over all trips, but the spread of trip level residuals is a concern. The large deviation of EC within trips from city lines can for a significant part be explained by the way in which SoC is rounded off as an integer value. As a robustness analysis requires a more reliable estimate than a single trip can provide, the validation is extended to the level of vehicle tasks.

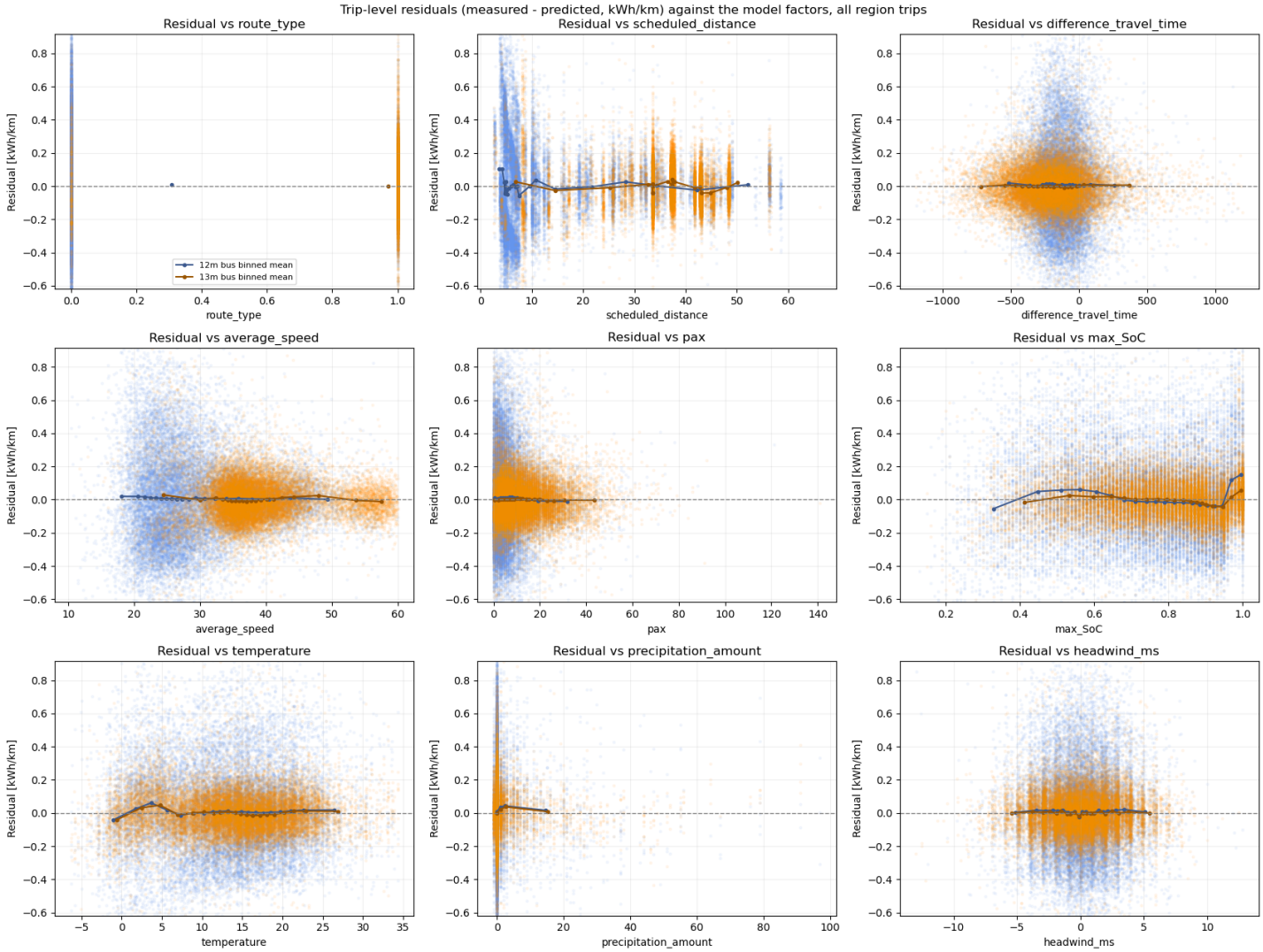


Figure 6.6: Trip level residuals plotted across the range of values of each factor

The short trips are thus the main source of the deviations that are visible in the figure above. A trip of three kilometres consumes about 3 kWh at an EC of 1 kWh/km, which is less than a single SoC step. The recorded EC of such a trip is therefore either zero or one full step, regardless of what the bus actually consumed, which translates into a very large error per kilometre. For long trips, this rounding error is spread over many kilometres and has little effect.

6.1.4. Validation on vehicle task level

To combat the problems that the data set gives on trip level, especially for short trips, the validation is extended to vehicle task level. Looking at a complete vehicle task should balance extreme values between trips. If a short trip has a high EC that is in part due to rounding of the SoC value, it is more likely that the next trip has a rounding error that works in the opposite direction. This is also the level that matters for this research, as the question whether a bus makes it to the next charger depends on the total EC of all trips in between.

The method of collecting coherent vehicle tasks is complicated due to unrecorded trips and bus swaps at the start of, or during the day. An analysis of this for the Apeldoorn and Lelystad depot is added to appendix A. The data shows how a large part of the vehicle tasks is performed by another vehicle type than the schedule prescribed. Furthermore, the schedule prescribes that each vehicle task should be handled by a single vehicle. In reality this was often not the case. This can for example happen when the SoC is depleting faster than expected and real-time adjustments in the form of bus swaps are necessary. The remaining trips are then covered by the new bus (or until another swap is initiated).

The solution is found in another trip grouping method. Instead of looking at purely predefined vehicle tasks, the validation is based on all trips driven by a unique vehicle on a single day. In that way it is still possible to balance out extreme values due to rounding errors on trip level, while having actual fleet schedules is not necessary. The residuals from all vehicle-day combinations are presented in figure 6.7. For the sake of simplicity, these vehicle-day combinations are in the remaining part of this thesis referred to as vehicle tasks.

The figure shows that the residuals are smaller and have a lower spread on vehicle task level, compared to trip level. Both the 12m and the 13m type show significant improvements. The 13m type now records 90% of all EC estimations within a deviation of about 0.1 kWh/km (the 5^{th} and 95^{th} percentile are -0.101 and $+0.096$). For the 12m bus type the improvement is even greater (the R^2 rises from 0.27 to 0.76), but it is still slightly less accurate, with the range between the 5^{th} and the 95^{th} percentile coming in at -0.118 and $+0.104$ respectively. The mean residual is close to zero for both types (-0.002 and $+0.000$ kWh/km), so the model is not biased at this level. A complete overview of the statistics, including a comparison with the validation on trip level, can be found in table 6.6.

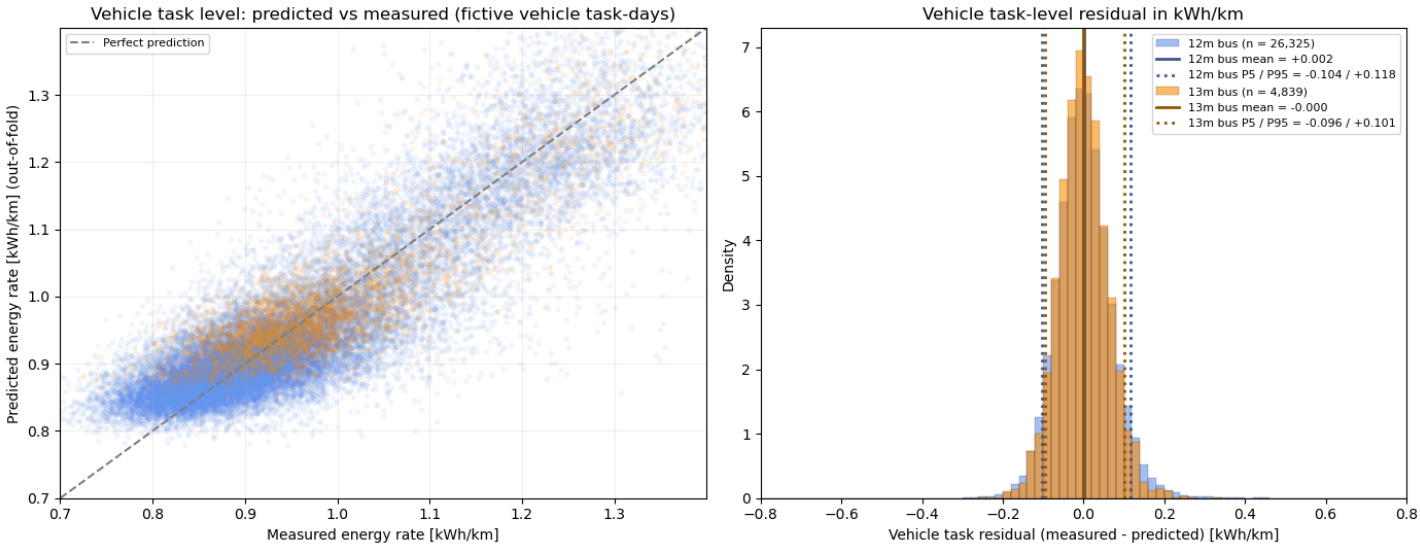


Figure 6.7: Scatter plot (left) and histogram (right) with the vehicle task level residuals

The performance of the model on vehicle task level comes close to the performance reported in the literature on EC estimation that was discussed in the literature review. A more extensive comparison, including the question to what extent the metrics can be compared across studies, is given in section 6.1.5. While an error of 0.1 kWh/km is more workable than the error on trip level, it implies that there is an unexplained uncertainty of roughly 10% of the EC, which also has to be accounted for in

Table 6.6: Model accuracy metrics on vehicle task level

Vehicle type	Level	R^2	RMSE (kWh/km)	MAE (kWh/km)	P5 residual (kWh/km)	P95 residual (kWh/km)
12m bus	Trip level	0.2660	0.2688	0.2011	-0.4970	0.3885
	Relative change [%]	185.7	-74.1	-73.7	76.4	-73.3
	Vehicle task level	0.7598	0.0697	0.0528	-0.1175	0.1037
13m bus	Trip level	0.4691	0.1357	0.0951	-0.2024	0.1860
	Relative change [%]	62.9	-54.9	-49.8	50.1	-48.1
	Vehicle task level	0.7643	0.0613	0.0478	-0.1010	0.0965

the simulation. This is done through a Monte Carlo simulation in the sensitivity analysis (section 6.4). Where the EC of a vehicle task is calculated deterministically in the discrete simulation, the sensitivity analysis adds a random error that is based on the residuals presented in this section.

6.1.5. Reflection on the EC model

Before the EC model is used in the simulation, it is placed in the context of the literature, as the performance of the model determines how far the outcomes of the sensitivity analysis can be trusted. A direct comparison with other studies is difficult for three reasons. First, the studies differ in granularity: the models with the lowest errors use high-frequency speed data, for example the MAPE of 2.8% per second (4.2% on trip level) of Chen et al. (2021), which was not available for a full year of trips in this study (appendix B). Second, the discrete SoC steps are a property of the data of the operator that no model can explain, especially for short trips. Third, the metrics differ. Many studies report the MAPE, which is sensitive to trips with a low EC, whereas this study reports the R^2, RMSE and MAE in kWh/km. With this in mind, the trip level accuracy of this study (a MAPE of about 22% for the 12m bus and 10% for the 13m bus) is lower than the 4.4% of Z. Zhang et al. (2024) and the 12–14% of Ji et al. (2022) and Li et al. (2021). On vehicle task level, the MAE of about 0.05 kWh/km is roughly 5% of the average EC, which is in the range of the better-performing studies. Note that all the studies that were reviewed report their accuracy per trip or at an even finer resolution. None of them validates the accuracy on the combined trips of a bus, which is the level at which a model is used for fleet scheduling. That makes the validation on vehicle task level a contribution in itself. It also calls for caution: an error of about 10% is of the same size as the redundancy that is studied in the sensitivity analysis, which is why the unexplained variance is explicitly included in the simulation.

The relationships that the model finds are in line with the literature. Temperature dictates the EC, as also found by Z. Zhang et al. (2024) and Ji et al. (2022). The optimum of about 19°C is lower than the 23.7°C of Ji et al. (2022), and the increase in EC at low temperatures is larger (48% between the lowest and highest EC within the 5^{th} to 95^{th} percentile for the 12m bus, against 19.8% at 0°C compared to the optimum). Differences in the heating system, the insulation of the bus and the range of temperatures in the data may explain this, but the direction and the quadratic shape agree. Note that very low temperatures are rare in the data (less than 5% below freezing), which is why the 5^{th} to 95^{th} percentile range is used to compare factors. The SoC at the start of a trip is the second most important factor, in line with Abdelaty and Mohamed (2021), although part of the effect is a measurement artefact (SoC drops). For pax the relationship is positive, which is the same direction as found in studies on vehicle weight, but pax based on check-ins cannot separate the added weight from the effect of more frequent stopping.

Several factors from the literature are not part of the model, which is partly the explanation for the unexplained variance. The road gradient is among the most influential factors in other studies: Chen et al. (2021) reduce the MAPE of their model from 7.0% to 4.9% by adding the road gradient and road class. Although the Netherlands is relatively flat, routes that cross the Veluwe have height differences of 80 to 100 metres. The scheduled distance, which turned out to carry line-specific information for the 13m bus, may be capturing a part of this effect. The number of stops and traffic lights (Li et al., 2021) is only indirectly present through the average speed and route type, and an occupancy rate from weighing systems or cameras would be a better indicator of weight than check-ins. Finally, driving behaviour is not included, as the indicator that was explored was too data-intensive (appendix B). The omission of

these factors is the consequence of opting for a model that is broadly applicable and works with data that operators already collect. An operator with a more varied terrain should add the gradient first.

6.1.6. Conclusion on the EC model

The EC model of each bus type is a degree two regression on the nine factors (equation 6.1 and table 6.1), which is the formula that is used to calculate the EC per kilometre of a trip in the simulation. The model identifies temperature as the most influential factor by a wide margin, followed by max SoC and pax. On trip level the model explains a limited share of the variance (R^2 of 0.27 for the 12m bus and 0.47 for the 13m bus), which is largely caused by the discrete SoC steps in the data. On vehicle task level, where these steps average out, the model explains 76% of the variance for both bus types, with 90% of the estimates within about 0.1 kWh/km of the measured value. The remaining uncertainty is not negligible and is therefore included in the sensitivity analysis through a Monte Carlo simulation (section 6.4).

An additional factor that was explored in an early stage of this research is a speed dynamics indicator, derived from the speed profiles of the buses. The analysis showed a promising improvement in EC estimation for a small sample of 125 trips, but the data requirements of the indicator are too high to apply it to a complete year of trips. The indicator is therefore not part of the EC model. The analysis is described in appendix B.

Conclusion on the EC model (SRQ1)

Temperature is by far the most influential factor for the EC of an electric bus, with a U-shaped relationship and an optimum around 19°C. At the coldest temperatures that occur regularly, the EC is 35% to 48% higher than at this optimum, which means that a bus has roughly a quarter to a third less range. Max SoC and pax follow at a large distance, whereas precipitation and the interaction effects hardly matter. The model is accurate enough to assess schedules on vehicle task level, but not on trip level. This has three implications for the rest of this study. First, temperature should be the main factor in the scenarios. Second, max SoC and pax have to be part of the simulation, and the SoC through the trips that came before. Third, the unexplained variance should be included in the outcomes of the simulation.

6.2. Scenario setup

Now that the relationship between each factor and EC is established, the next step in preparation of the robustness simulation is to define the scenarios. To make a scenario relevant for a case study, two things are important. First, does the scenario have a meaningful effect on EC, and second, does the scenario occur frequently enough to take its effects seriously? Temperature proved to be the most influential factor for EC by quite a margin, and thus should be the main factor to vary between different scenarios. There is a steep increase in EC when temperatures get lower than 5°C. With extrapolation one can assume that the same holds on the other side of the parabola, but because temperatures higher than 35°C rarely occurred in the dataset, the focus is on low temperatures. The following sections first derive the probability of the temperature scenarios (section 6.2.1) and then the number of passengers (section 6.2.2). After that, the factors that are left out of the scenarios, or are simplified, are discussed (section 6.2.3) before the final set of scenarios is presented (section 6.2.5).

6.2.1. Future weather statistics

Meaningful statistics about weather cannot be derived from a single year. The KNMI itself looks at a period of 30 years to define a climate and its characteristics. A similar approach is taken in this study. However, as IPCC (2021) indicate, the climate is changing. Therefore, historical data may not fully capture future scenarios. Luckily, van Dorland et al. (2023) already looked into this question. Their work models the effects of climate change in a Dutch context. With the data available on the KNMI website, both historical temperature data and the expected future values can be analysed.

The daily temperature during the period 1991–2020 is gathered from various weather stations throughout the case study area. With the tool that the KNMI provides, the temperature for the same period, but as if it was in the expected climate of 2050, can be found as well. The 2050 scenario assumes that moderate action is taken to combat climate change. Combining the datasets of both the climate for 2005 (the period 1991–2020) and 2050 results in an expected climate for 2027.

On a smaller time scale, temperature fluctuates as well. Within a day, temperatures usually reach their lowest point around sunrise and peak in the afternoon. However, the pattern is influenced by many weather aspects and often deviates from the theoretical curve. van Engelen and Geurts (1983) point out that the difference between the lowest and highest temperature of a day is on average lower during the colder months, compared to the summer period. These cold months are also the most interesting period for this research. To prevent overanalysing the daily temperature patterns, it is assumed that every scenario has a single temperature for the whole day.

As stated before, the temperature range in which high EC rates occur lies between -5°C and 5°C. This is translated into three scenarios: -5°C, 0°C and 5°C, which are used as input for the EC calculation during the simulation. To derive the probability of these temperatures occurring, the values are transformed into sub-ranges. The following values are captured within each sub-range: $\mathrm{Temp} \leq -2.5°\mathrm{C}$, $-2.5°\mathrm{C} < \mathrm{Temp} \leq 2.5°\mathrm{C}$ and $2.5°\mathrm{C} < \mathrm{Temp} \leq 7.5°\mathrm{C}$. In figure 6.8 the normalised histogram for the daily temperature during the period of 30 years centred around 2027 is shown. The corresponding probabilities for the ranges are 1.94%, 8.91% and 22.68% respectively. Given that there are on average 200 weekdays, school holidays excluded, on which the normal weekday schedule operates, this translates to 3.88 days for the -5°C scenario, 17.81 days for 0°C and 45.36 days for 5°C. The remaining 133 days are warmer than 7.5°C. As these are not expected to lead to problems with a schedule that is designed for lower temperatures, they are not simulated.

The focus on low temperatures does not mean that high temperatures cannot be problematic. The parabola of the temperature suggests that the EC also increases at high temperatures, and other studies report a higher EC when the air conditioning is used extensively (Belloni et al., 2025; Ji et al., 2022). In the data of this study, however, only a few trips took place above 30°C, so the shape of the curve in that range is uncertain. More importantly, high temperatures occur less frequently in the Netherlands than low temperatures. For regions with a warm climate, the scenarios should therefore be extended to high temperatures.

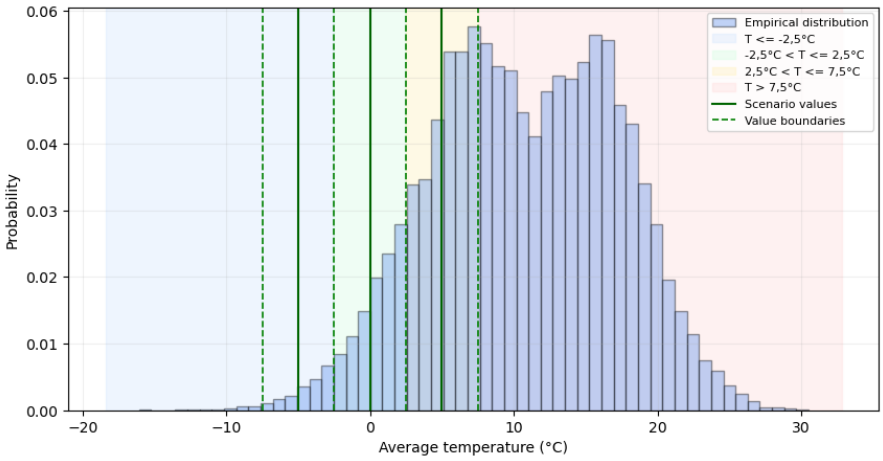


Figure 6.8: Normalised histogram of the daily mean temperature for the expected climate of 2027, based on the period 1991–2020 and the KNMI climate scenario for 2050

With the temperature scenarios defined, the second factor that varies between scenarios is the number of passengers.

6.2.2. Pax generation

The number of passengers on a trip is relevant for two reasons. It has a mild influence on EC, and, more importantly, it determines the societal cost of a cancelled trip, as the delay that passengers experience is multiplied by the number of passengers on that trip. However, the number of pax fluctuates between days, lines and unique trips. To get representative pax values for the scenarios, three steps are taken: the missing pax values are estimated, the days that can be used to represent a normal weekday are selected, and the daily totals on those days are translated to pax levels for the scenarios.

Estimating missing pax values. Pax is not recorded for all trips, for example because trips were not recorded or were filtered out. These missing values are estimated using the intra-day and intra-line correlations between measured trips. Note that this analysis is done for weekdays only, because the simulation also uses the schedule for weekdays. The intra-day correlation shows how the total number of pax on a given day relates to other days. If a morning trip has more pax than average, it can be assumed that those pax also have to return later in the day, so the afternoon trip should also have more pax. Intra-line correlation refers to the relationship between all trips across a line. For example, if trip 1001 has 5 pax, trip 1002 has 1 pax and trip 1003 has 7 pax on average, it can be assumed that trip 1002 has a high chance of getting lower pax on any given day, compared to the other trips. This can for example be caused by rush hour traffic where people flock from their homes to go to work: a trip in one direction serves a high number of passengers, but the trip the other way around at the same time is almost empty. Combining this information leads to an estimate that consists of three parts. The base is the average pax of the trip. This is adjusted with a line-day effect: the ratio between the average of the measured trips of that line on that day and the line average. Finally, a random residual is added, which is drawn from the leave-one-out validation of the same line. Measured values are never replaced.

Selecting representative days. The estimates are only meaningful if enough of the day was measured. A day is therefore included for a depot if at least 80% of the vehicle tasks of the depot could be traced in the data, and at most 40% of the service trips had to be estimated. This results in 225 usable weekdays for Apeldoorn and 210 for Lelystad, out of roughly 250 weekdays in 2025. Of these, 185 and 171 respectively are school days (out of about 200 school days per year), and the remaining 40 and 39 are days in school holidays. The days in school holidays are labelled separately, as the demand is considerably lower on those days and the benchmark schedule describes the normal school day. Note that these numbers only describe the sample on which the distribution of the daily pax is based. They are not used to determine the number of operating days per year, for which the 200 weekdays outside the school holidays (see section 6.2.1) are used. The holidays are derived from the 2025 calendar for region Midden (Apeldoorn) and region Noord (Lelystad).

Daily totals and temperature. The distribution of the estimated daily total pax on these days is shown in figure 6.9. School days are split in cold days (daily mean temperature of 7.5°C or lower) and warm days. The colder days are the days that are relevant for the scenarios, as all scenario temperatures lie below 7.5°C. This split is also supported by the relationship between the temperature and the number of pax, shown in figure 6.10. There is a visible, moderate negative correlation between temperature and the daily total of pax on school days (Pearson $r = -0.49$ for Apeldoorn and -0.42 for Lelystad). On average, the number of pax increases by 94 pax per day for every degree that the temperature decreases in Apeldoorn, and by 20.5 pax per day in Lelystad. This means that the days with the highest EC are also the days with more passengers, which is the combination that is the most harmful in case of cancellations. Note that the correlation is a simplification, as a single temperature explains only part of the variation (R^2 of 0.24 and 0.18).

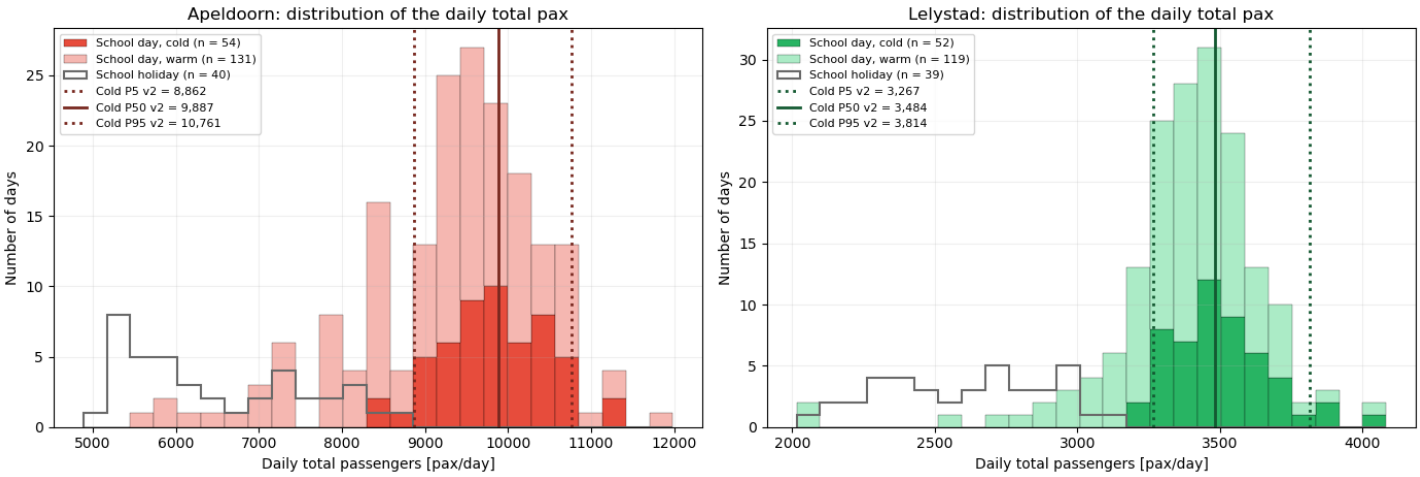


Figure 6.9: Distribution of the estimated daily total pax for the Apeldoorn (left) and Lelystad (right) depot. The dotted lines indicate the 5^{th} and 95^{th} percentile of cold school days, the solid line the median

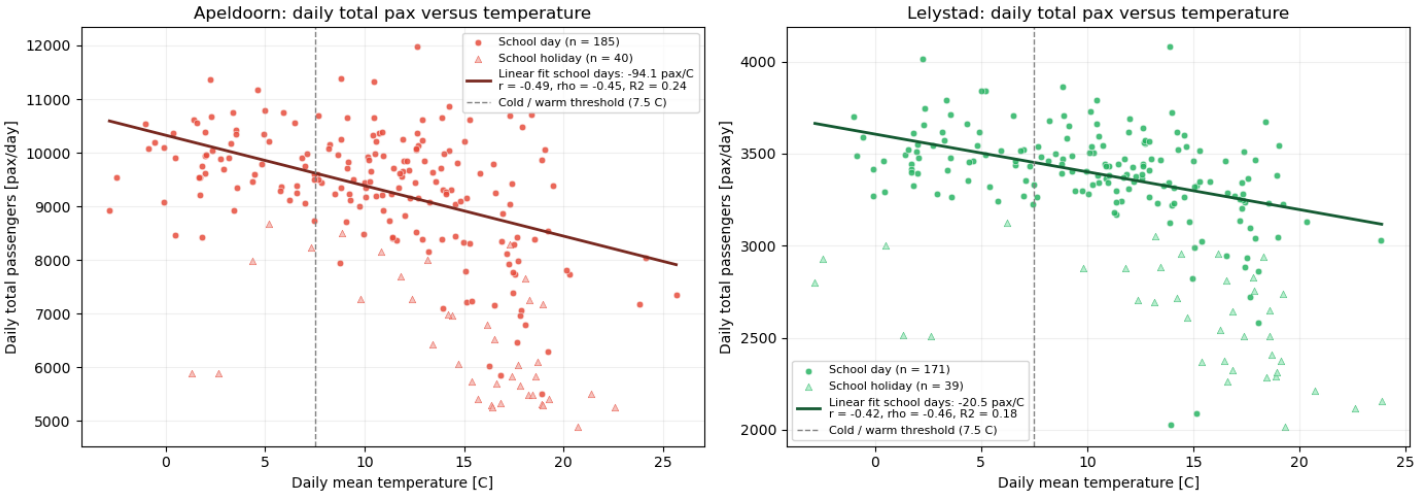


Figure 6.10: Estimated daily total pax against the daily mean temperature for Apeldoorn (left) and Lelystad (right)

Pax levels for the scenarios. Rather than using the exact correlation, the spread in the daily totals on cold school days is used to give an indication of the consequences for passengers on a busy and a less busy day. Three levels are defined: the 5^{th} percentile (P5), the median (P50) and the 95^{th} percentile (P95) of the daily total on cold school days. For Apeldoorn these are 8,862, 9,887 and 10,761 pax per day, for Lelystad 3,267, 3,484 and 3,814 pax per day. The pax of each individual trip is derived from the days that lie closest to the percentile, and scaled such that the sum over the day matches

the level. The three levels are combined with each of the three temperature scenarios. The levels represent the lower tail, the middle and the upper tail of the distribution. As values around the 5^{th} and 95^{th} percentile occur less often than values around the median, the levels are not weighted equally, but with 20%, 60% and 20% respectively. As the variation in pax within cold days is much smaller than the effect of temperature, pax is treated as independent of the temperature scenario once the cold days are selected. In this way, the combination of temperature and pax leads to nine scenarios per depot, which also provides an answer to the second sub-research question. How these scenarios relate to the factors that are not varied is discussed next.

6.2.3. Omissions and simplifications

The results of section 6.1 are used to decide which factors should be varied in the scenarios, which can be derived from the simulation itself and which can be left out. Factors are only worth including when they have a notable influence on the model, as measured by the drop in R^2 when the factor is dropped (figure 6.2), and when this fits in a practical sense in the simulation. For each of the nine factors it is explained how it is handled. In the simulation all factors that are not varied are set to their mean value in the training data.

Temperature is by far the most important factor and is therefore the main factor that is varied between the scenarios. **Max SoC** is next in line when it comes to importance. This is included in the simulation itself, rather than in the scenarios, because the initial SoC of a trip depends on the previous trips and charging moments of the vehicle task. The EC calculation within the simulation directly derives the SoC at the start of a trip, which is then used to calculate the EC of that trip. For **average speed** a similar method is applied. While in reality the average speed differs between trips, in the simulation it is assumed that the trip time does not differ from the scheduled trip time. The average speed is therefore calculated by dividing the scheduled distance by the scheduled trip duration. This does not capture the day-to-day variation in speed, but the importance of the factor is low, and the variation that remains is part of the random error that is added in the sensitivity analysis. The **scheduled distance** is directly available from the schedule.

The third most influential factor is **pax**. The influence on EC is less dramatic compared to temperature, but knowing the number of passengers on a trip is also relevant for estimating the societal cost of delays as a result of trip cancellations. Moreover, pax on the same line and trip number tend to differ between days. When fewer people travel with public transport, the impact of a cancellation is also lower, and vice versa. This is why pax is included as the second variable in the scenarios, as described in the previous section.

The remaining factors are not varied. **Headwind** is not taken into account, as the case study features hub-and-spoke networks. Most trips in one direction are followed by a return trip in the opposite direction, so the effect of wind largely cancels out between the two trips. Besides that, the drop in R^2 is small for the 12m bus. The **difference in travel time** has a similar importance as average speed and headwind. As the simulation uses the scheduled travel time, it is set to its mean value. **Route type** and **precipitation** rank lowest. Route type is a fixed characteristic of a trip that is already captured by the scheduled distance and average speed with which it correlates. Precipitation has almost no impact on the model and is therefore not included.

With these choices, the simulation uses a simplified version of the EC model. Whether this has consequences for the performance of the model is tested in the next section.

6.2.4. Effect of simplifications

The exclusion of some of the factors and the simplifications in the handling of the factors that are included may cause the performance of the model to deteriorate. To test whether this is the case, the performance of the model on vehicle task level under these simulation adjustments is compared to the normal EC model. The comparison is made on the vehicle tasks of the schedules of both depots (allowing for a maximum of two bus changes per task), for each depot and bus type, and is shown in figure 6.11.

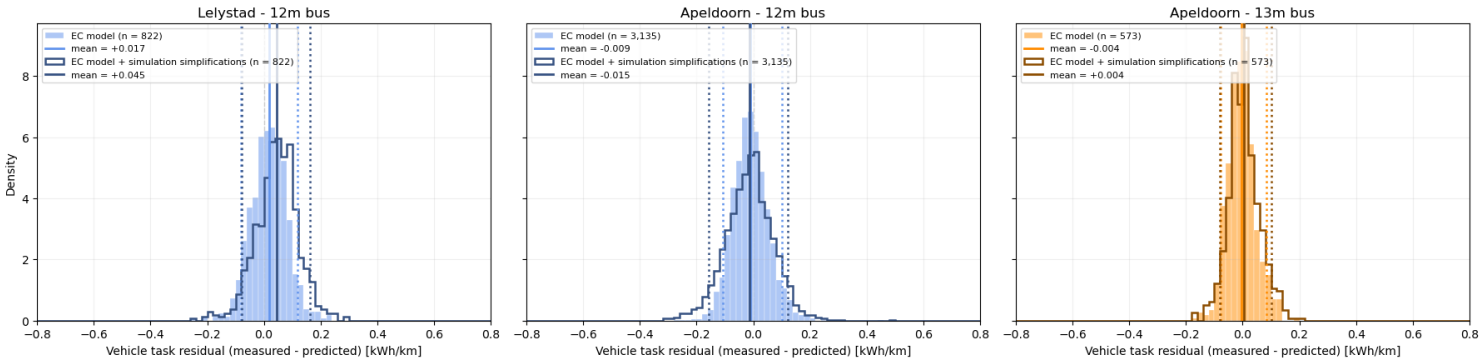


Figure 6.11: Residuals for vehicle tasks from the Apeldoorn and Lelystad depot calculated for both the base EC model and the EC model with simplifications for the simulation

The simplifications shift the average residual only slightly. For the 12m bus in Lelystad the mean residual changes from -0.023 to -0.045 kWh/km, which means that the simulation underestimates the EC by roughly 4–5%. In Apeldoorn the mean residual of the 12m bus changes from $+0.012$ to $+0.027$ kWh/km, an overestimation of about 3%, and for the 13m bus it is nearly unchanged ($+0.008$ to $+0.003$). The spread of the residuals increases a little for the 12m bus, most visibly in Apeldoorn. The effect of the simplifications is therefore small compared to the unexplained variance of the model itself. A noteworthy consequence is that the simulation is slightly optimistic for Lelystad and slightly conservative for Apeldoorn. This should be kept in mind when the results of both depots are compared.

6.2.5. Set of scenarios

The temperature scenarios and the pax levels together form the set of scenarios that is used in the simulation. Table 6.7 gives an overview. Every temperature is combined with every pax level, so there are nine scenarios per depot. The weight of a scenario is the product of the probability of the temperature scenario and the weight of the pax level. This leads to the expected number of days per year per scenario, which is the number by which the outcomes of a scenario are multiplied to arrive at annual numbers. The scenarios with an average number of pax at 5°C are the most common, with 27.22 days per year. The most extreme scenarios, at -5°C, together make up only 3.88 days a year.

The general assumptions in the simulation are the same for all scenarios. A day starts at 04:00 with all buses fully charged (100% SoC). The EC of each trip is calculated with the simplified EC model, using the temperature and pax of the scenario, and the SoC and average speed of the trip. A bus may not drop below the minimum SoC, and in this case buses are redirected to a charger (see section 3.3.1). The scenario affects only the EC and the societal cost of cancellations: the schedule, the chargers and the charging speed are the same in all scenarios.

Table 6.7: Overview of the scenarios. Every temperature is combined with every pax level, which results in nine scenarios per depot. The expected number of days per year of a scenario (D_s) follows from the 200 operating days, the probability of the temperature and the weight of the pax level

Temperature	**Probability (%)**	**Days per year (total)**	**Per pax level (P5 / P50 / P95)**
-5°C	1.94	3.88	0.78 / 2.33 / 0.78
0°C	8.91	17.81	3.56 / 10.69 / 3.56
5°C	22.68	45.36	9.07 / 27.22 / 9.07
Pax level	**Weight (%)**	**Apeldoorn (pax per day)**	**Lelystad (pax per day)**
5th percentile (P5)	20	8,862	3,267
Median (P50)	60	9,887	3,484
95th percentile (P95)	20	10,761	3,814

6.2.6. Reflection on the scenarios

The scenarios are based on an average year. If low temperatures are distributed in a skewed way over the years, for example one very cold year after four mild years, a schedule that is based on an average year would be too robust in the mild years and still lack robustness in the cold year. As the consequences of such a year are felt directly in the performance of a concession, this should be considered when the results are used. The scenarios also stop at 5°C. Days that are warmer than that are not simulated, so the outcomes say nothing about the cost of cancellations on those days, nor about the problems that high temperatures could cause through air conditioning. The daily total of pax is also a simplification. It is based on estimates for trips that were not recorded, and the pax levels (5^{th}, 50^{th} and 95^{th} percentile with fixed weights of 20/60/20) are a discrete representation of a continuous distribution. The weights are an assumption, with less weight on the extremes as they occur less often. They are meant to show the sensitivity of the societal cost to the number of pax, rather than to give a precise prediction.

Conclusion on the scenarios (SRQ2)

Temperature and pax are the two factors that vary between the scenarios. Three temperatures (-5°C, 0°C and 5°C) that together occur on about 67 days per year are combined with three pax levels, which results in nine scenarios per depot with an expected number of days per year. Max SoC and average speed are calculated within the simulation, while headwind and precipitation are left out. The effect of these simplifications on the EC estimate is small compared to the unexplained variance of the model. The scenarios are based on an average year and a daily mean temperature, which should be kept in mind when the outcomes are used.

6.3. Simulation of the benchmark schedule

The first analysis with the freshly developed scenarios is to assess whether the simulation works as intended, and to see how the benchmark schedule (the actual schedule operated in 2025 on normal weekdays) performs. The schedules of Apeldoorn and Lelystad are simulated separately, meaning that the depots do not share any resources. This would be different if the case study included a multi-depot setup.

The story is built up in two steps. First, the results of individual scenarios for Lelystad are discussed in detail. For this part the residuals of the EC model are not yet taken into account, making it a discrete simulation: every scenario results in a single outcome. This shows how low temperatures affect trip completion and how individual cancelled trips lead to costs. Then the view is zoomed out, to see how these costs add up to the outcomes of the benchmark schedule of Lelystad and Apeldoorn. For this second step the unexplained variance is included through the Monte Carlo simulation, so the outcomes are the mean over 1,000 runs. These outcomes are what the alternative schedules are compared with in the final section.

6.3.1. Individual scenario performance for Lelystad

The results for two scenarios are shown in figures 6.12 and 6.13, together with tables 6.8 and 6.9. For the scenarios that have a temperature of 5°C, no problems occur. The first trips must be cancelled when the temperature drops to 0°C and this trend accelerates as the temperature reaches -5°C. Because EC is largely impacted by temperature, and the different levels of pax are mainly included to show the effects on societal cost under different occupancy rates, only the scenarios with average pax are shown.

The visualisation of events in the simulation makes a distinction between planned and unplanned events. Events that are completed as planned are shown as half transparent. These can be either service trips (green), dead-heading trips (grey) or charging events (blue). When the simulation has to cancel trips to go to a charger, the cancelled service trips are coloured red and the cancelled dead-heading trips are dark grey. They are pictured in the bottom half of the vehicle task row. The upper half of the row indicates how the simulation solved the situation. An orange event indicates an unplanned dead-heading trip towards a charger and a dark blue event indicates an unplanned charging event. This visualisation setup is used throughout the rest of this chapter. In addition to the visualisations, details of the cancelled trips are given in table format.

The outcomes of the first scenario under investigation, with a temperature of 0°C, are displayed in figure 6.12 and table 6.8. They show how the simulation under these conditions leads to six cancelled service trips, which are all in the evening. Vehicle task 4503 has to divert to charging earlier than planned, and therefore misses two service trips just after 20:00. Vehicle task 4510 faces a problem at the end of its duty. There, four service trips between 23:10 and 24:10 are cancelled. In the case of Lelystad the frequency of the city lines is halved during the evening. This is done to reflect lower demand in those hours. While the number of travellers that are affected is often lower, as late trips tend to have fewer passengers, the impact for each passenger is considerable if that trip is the last of the day for that line. This effect is visible at the cancelled trips in vehicle task 4510. Although the expected number of pax is about the same as for the other trips, the longer headway of 48 minutes leads to a higher passenger delay per trip. Note that for the very last trip of a line the headway should be considered as higher in reality, as there is no next bus on that day. This effect is not captured by the headway assumption.

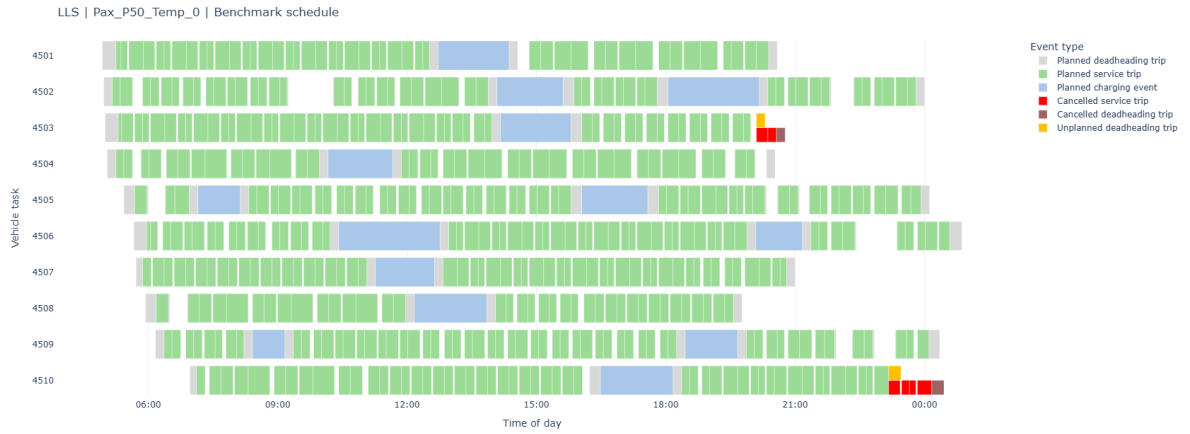


Figure 6.12: Overview of trip completion for all Lelystad vehicle tasks under the scenario with average pax and a day temperature of 0°C

Table 6.8: Cancelled trips for the benchmark schedule of the Lelystad depot under the scenario with average pax and a temperature of 0°C

Vehicle task	Trip	Route	Dir.	Planned dep. / arr.	Planned (km)	Headway (min)	Expected pax	Total pass. delay (hrs)	Calculated SoC (%) dep. / arr.
4503	1057	4916	1	20:06 / 20:22	6.25	15.0	4.33	1.08	12.26 / 9.03
4503	1060	4916	2	20:22 / 20:34	3.69	15.0	2.99	0.75	9.03 / 7.03
4510	1066	4901	2	23:10 / 23:26	6.26	48.0	2.08	1.67	11.81 / 8.63
4510	1069	4904	1	23:28 / 23:39	5.05	48.0	3.33	2.66	8.63 / 6.04
4510	1070	4904	2	23:39 / 23:48	2.70	48.0	1.70	1.36	6.04 / 4.55
4510	1067	4901	1	23:50 / 24:10	8.64	48.0	2.64	2.11	4.55 / -0.03

Figure 6.13 shows how the problems evolve when temperatures drop further to -5°C. Only three of the ten vehicle tasks remain unaffected, and the number of cancelled service trips increases to 39 (table 6.9). Focussing again on vehicle task 4503, the vehicle task faces problems earlier in the day compared to the previous figure. This is a logical consequence of SoC draining faster due to higher EC and confirms the logic of the simulation.

Here, the difference between trips during peak hours and off-peak hours is visible in terms of pax. That validates the ability of the simulation to weigh passenger impact throughout the day and also implying that the cancellation of trips should be avoided at those times in particular. Another thing that stands out from the table is that trips with a relatively high calculated SoC can be cancelled. These are the trips that are cancelled because the bus cannot reach the trip in time, while it still has to travel from the charger back to the next trip's starting point. In all other trips, the simulation follows a logic where trips are cancelled because the bus's SoC is projected to dip under the 5%

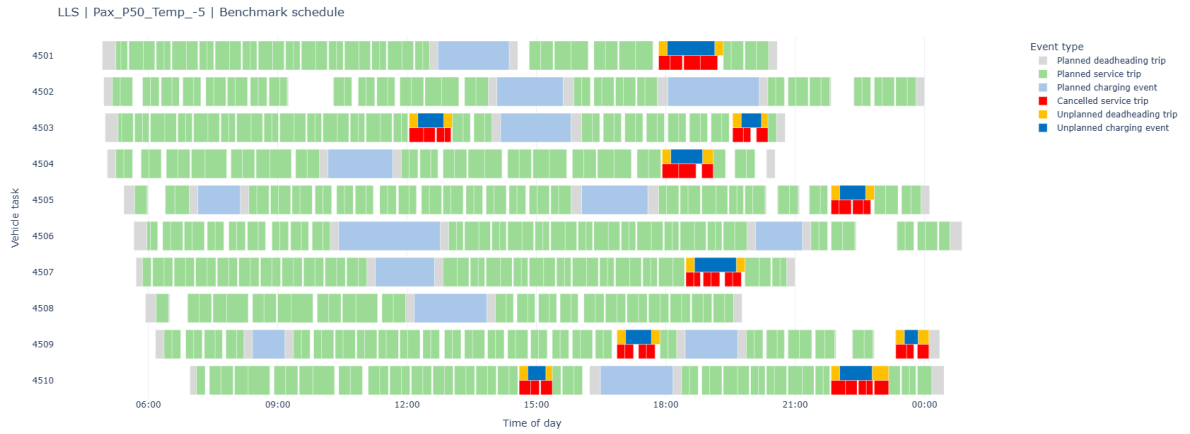


Figure 6.13: Overview of trip completion for all Lelystad vehicle tasks under the scenario with average pax and a day temperature of -5°C

Table 6.9: Cancelled trips for the benchmark schedule of the Lelystad depot under the scenario with average pax and a temperature of -5°C

Vehicle task	Trip	Route	Dir.	Planned dep. / arr.	Planned (km)	Headway (min)	Expected pax	Total pass. delay (hrs)	Calculated SoC (%) dep. / arr.
4501	1047	4902	1	17:50 / 18:06	6.17	18.0	12.68	3.81	14.56 / 10.70
4501	1052	4902	2	18:06 / 18:23	6.20	18.0	2.44	0.73	10.70 / 7.00
4501	1039	4907	1	18:25 / 18:48	12.96	20.0	4.84	1.62	7.00 / -0.42
4501	1042	4907	2	18:48 / 19:12	11.13	20.0	3.83	1.28	-0.42 / -7.35
4503	1029	4901	1	12:03 / 12:23	8.64	24.0	4.57	1.83	15.68 / 10.74
4503	1030	4901	2	12:23 / 12:39	6.26	24.0	4.39	1.76	10.74 / 6.98
4503	1031	4904	1	12:41 / 12:52	5.05	24.0	8.21	3.28	6.98 / 3.85
4503	1032	4904	2	12:52 / 13:01	2.70	24.0	6.45	2.58	29.86 / 28.33
4503	1055	4911	1	19:33 / 19:48	6.84	15.0	3.25	0.81	13.69 / 9.80
4503	1056	4911	2	19:48 / 19:58	4.65	15.0	1.85	0.46	9.80 / 7.12
4503	1057	4916	1	20:06 / 20:22	6.25	15.0	4.33	1.08	7.12 / 3.28
4504	1037	4907	1	17:55 / 18:18	12.96	20.0	3.95	1.32	9.73 / 2.50
4504	1040	4907	2	18:18 / 18:42	11.13	20.0	1.86	0.62	2.50 / -4.18
4504	1051	4902	1	18:50 / 19:06	6.17	18.0	7.70	2.31	-4.18 / -8.40
4505	1065	4906	1	21:50 / 22:01	4.45	30.0	7.45	3.73	13.79 / 11.11
4505	1066	4906	2	22:01 / 22:18	5.38	30.0	2.32	1.16	11.11 / 7.84
4505	1061	4911	1	22:20 / 22:35	6.84	30.0	3.53	1.76	7.84 / 3.79
4505	1062	4911	2	22:35 / 22:45	4.65	30.0	2.62	1.31	3.79 / 0.99
4507	1055	4904	1	18:28 / 18:39	5.05	24.0	8.22	3.29	11.68 / 8.65
4507	1056	4904	2	18:39 / 18:48	2.70	24.0	1.76	0.70	8.65 / 6.97
4507	1055	4903	1	18:52 / 19:03	4.98	18.0	9.53	2.86	6.97 / 3.85
4507	1056	4903	2	19:03 / 19:15	5.20	18.0	5.15	1.55	3.85 / 0.61
4507	1057	4903	1	19:22 / 19:33	4.98	18.0	6.17	1.85	0.61 / -2.56
4507	1058	4903	2	19:33 / 19:45	5.20	18.0	5.21	1.56	-2.56 / -5.95
4509	1047	4903	1	16:52 / 17:03	4.98	18.0	11.40	3.42	10.70 / 7.61
4509	1048	4903	2	17:03 / 17:15	5.20	18.0	17.08	5.13	7.61 / 4.16
4509	1049	4903	1	17:22 / 17:33	4.98	18.0	10.79	3.24	4.16 / 0.95
4509	1050	4903	2	17:33 / 17:45	5.20	18.0	12.25	3.67	0.95 / -2.54
4509	1063	4911	1	23:20 / 23:35	6.84	30.0	2.51	1.25	14.14 / 10.29
4509	1064	4911	2	23:35 / 23:45	4.65	30.0	1.59	0.80	10.29 / 7.62
4509	1063	4902	1	23:50 / 24:06	6.17	36.0	2.54	1.52	7.62 / 3.89
4510	1035	4916	1	14:36 / 14:52	6.25	15.0	12.65	3.16	14.36 / 10.46
4510	1038	4916	2	14:52 / 15:04	3.69	15.0	8.04	2.01	10.46 / 8.09
4510	1037	4916	1	15:06 / 15:22	6.25	15.0	12.49	3.12	8.09 / 4.03
4510	1063	4901	1	21:50 / 22:10	8.64	48.0	4.23	3.38	13.82 / 8.84
4510	1064	4901	2	22:10 / 22:26	6.26	48.0	2.15	1.72	8.84 / 5.10
4510	1067	4904	1	22:28 / 22:39	5.05	48.0	3.56	2.85	5.10 / 2.06
4510	1068	4904	2	22:39 / 22:48	2.70	48.0	1.73	1.38	2.06 / 0.30
4510	1065	4901	1	22:50 / 23:10	8.64	48.0	4.14	3.31	32.59 / 28.09

The table shows the details of the route, The Vehicle task, Trip, Route and Dir. (direction) columns state general trip information. The headway (based on the outcomes of the qualitative network assessment) is multiplied with the expected pax to generate total passenger delay. This is the total amount of time that passenger face extra out-of-vehicle travel time and arrive late at their destination. The calculated SoC shows the projected SoC values, would the bus not be redirected to charging.

The algorithm cancels a trip if the expected SoC at arrival is lower than the minimum SoC of 5%, or if the bus would no longer be able to reach a charger from the destination of the trip above that minimum. That is why the first trip in a group of cancelled trips can still have a calculated SoC at arrival that lies above 5%: the remaining energy is needed to drive the dead-heading trip to the depot. The following trips in the group have lower calculated SoC values, and for some of them the SoC would even have dropped below zero. The sum of the cancelled trips determines the reliability of the scenario, and the delay and costs that were discussed above.

6.3.2. Benchmark schedule performance

The section above has shown how the simulation leads to results for individual scenarios. Table 6.10 summarises the outcomes of all nine scenarios for the benchmark schedule of both depots, per scenario day. The values are the mean over the 1,000 Monte Carlo runs. The first scenario columns describe the consequences for passengers: the share of completed service trips (completion rate), the total delay that passengers experience and the societal cost of that delay. The last column describes the consequence for the operator: the cost of cancelled trips, which consists of the missed income and the fines, minus the variable cost that is saved by not driving the cancelled trips.

Table 6.10: Performance of the benchmark schedule per scenario day (mean over 1,000 Monte Carlo runs), for Apeldoorn (red rows) and Lelystad (green rows). The scenario is described by the temperature and the pax level

Scenario	Days per year	Completion rate (%)	Pax delay (h/day)	Pax delay cost (€/day)	Operator canc. cost (€/day)
Apeldoorn					
-5°C, P5	0.78	96.42	155.1	5,037	2,640
-5°C, P50	2.33	96.30	186.1	6,044	2,751
-5°C, P95	0.78	96.20	207.3	6,732	2,838
0°C, P5	3.56	99.44	15.3	496	353
0°C, P50	10.69	99.39	22.3	725	403
0°C, P95	3.56	99.35	25.6	831	435
5°C, P5	9.07	99.99	0.4	13	8
5°C, P50	27.22	99.98	0.8	25	11
5°C, P95	9.07	99.98	0.9	28	12
Lelystad					
-5°C, P5	0.78	89.90	106.3	3,452	1,961
-5°C, P50	2.33	89.69	120.2	3,902	2,003
-5°C, P95	0.78	89.52	139.4	4,527	2,037
0°C, P5	3.56	96.63	31.6	1,026	649
0°C, P50	10.69	96.39	36.8	1,194	696
0°C, P95	3.56	96.19	43.4	1,409	734
5°C, P5	9.07	99.77	2.0	66	43
5°C, P50	27.22	99.72	2.7	87	53
5°C, P95	9.07	99.67	3.4	110	63

The table shows a clear pattern. At 5°C almost all trips are completed (99.7% to 100%), regardless of the number of pax. Compared to the discrete simulation, a few trips do get cancelled in some runs, as a vehicle task that is just feasible in the discrete calculation can fail when the EC is a bit higher than the model predicts. At 0°C the completion rate drops to about 99.4% in Apeldoorn and 96.4% in Lelystad, and at -5°C to about 96.3% and 89.7%. In Lelystad this means that roughly one in every ten service trips is cancelled on the coldest days. For the passengers these are the days on which public transport is clearly unreliable, and when this happens regularly, public transport is valued less. The passenger delay increases much faster than the completion rate decreases, as a share of the cancelled trips are trips with long headways in the evening. The number of pax has hardly any influence on the

completion rate, as it only has a mild effect on EC, but it does have a considerable effect on the societal cost: between the 5^{th} and the 95^{th} percentile of pax, the delay cost at -5°C differ by about 34% in Apeldoorn and 31% in Lelystad.

Apeldoorn has more than double the number of trips of Lelystad, of which also regional lines with more passengers, so the absolute passenger delay on the coldest days is higher (155–207 hours per day, against 106–139 hours). Relative to the size of the schedule, however, Apeldoorn performs better than Lelystad. This reflects the difference in charging infrastructure. When a bus in Apeldoorn needs to charge earlier than planned, it can do so at the pantograph chargers at the hub, so it does not need a long dead-heading trip and it can continue with a trip that starts at the same hub. In Lelystad the bus has to drive to the depot, about six kilometres from the hub, and charges slower, so it is absent for much longer.

For the operator, the cost of a cancelled trip consists of the missed income for the service hour that is not driven (€84 per hour) and the fine (€20 per trip), minus the variable cost that is saved because the trip is not driven. The latter is small. Cancelling trips therefore leads to a net loss for the operator, as all other cost components, such as the cost of the bus and the driver, stay the same. The share of the fines is higher in Lelystad because the average trip is shorter, and a trip therefore represents less income, whereas the fine is a fixed amount per trip.

The daily outcomes are translated to annual numbers by multiplying them with the expected number of days on which each scenario occurs (table 6.7) and summing over the scenarios. Table 6.11 shows the resulting annual outcomes of the benchmark schedules.

Table 6.11: Annual performance of the benchmark schedules, weighted over all scenarios by the expected number of days per year

	Apeldoorn	**Lelystad**
Completion rate at 5°C (%)	99.98	99.72
Completion rate at 0°C (%)	99.39	96.40
Completion rate at -5°C (%)	96.31	89.70
Expected completion rate (%)	99.61	98.26
Passenger delay (h per year)	1,131	1,253
Passenger delay cost (€ per year)	36,728	40,684
Operator cancellation cost (€ per year)	18,257	22,534

It is important to be precise about the definition of these numbers. They are not the total annual cost of cancellations, but the cost associated with the nine scenarios that are considered here: days with a mean temperature of 7.5°C or lower, and a normal school day pattern. Cancellations that occur outside these scenarios, for example on warmer days, are not part of the number. The same definition is used for all annual numbers in the remainder of this chapter. It is remarkable that the annual passenger delay of both depots is similar, while Lelystad has 35 vehicle tasks fewer than Apeldoorn. This shows the benefit of opportunity charging at the hub: the passengers of Apeldoorn are exposed to many more trips, but their delay does not increase proportionally. A bus that needs energy earlier than planned can charge during a layover instead of driving to the depot, and after a disturbance it is back in service sooner. The extreme conditions also dominate the annual numbers. The scenarios at -5°C occur on less than 2% of the days, but they are responsible for about 63% of the passenger delay cost in Apeldoorn and 38% in Lelystad.

6.3.3. Reflection on the benchmark results

The outcomes above depend on the decision rule of the simulation. A trip is cancelled as soon as the SoC would fall below the minimum, whereas in practice a bus with a low SoC triggers a warning on the dashboard, after which a driver contacts the control room and the bus can, for example, be swapped at a favourable moment by an emergency service. As this is difficult to model, it is left out of the simulation, which means that trips are often not cancelled in reality. At the same time, swapping buses takes time of the staff of the emergency service and cannot be done without limits. The simulated cancellations should therefore be seen as the trips that are at risk when no such measures are available, rather than as a prediction of the number of trips that really do not run, and the outcomes may be somewhat pessimistic. On the other hand, the simulation assumes that all buses start the day fully charged and

that chargers work at their nominal power, which could be optimistic. The net effect has not been quantified. The simulation also does not include variation between the trips within a vehicle task, and the simplifications of the EC model lead to a slight underestimation of the EC in Lelystad (about 4–5%) and an overestimation in Apeldoorn (about 3%), so the outcomes of Lelystad are probably slightly optimistic.

The outcomes also say something about the way in which robustness should be assessed. The expected completion rates of 99.61% and 98.26% look high, but they hide that the schedules have a weak spot on a few days per year: on the coldest days about one in ten service trips in Lelystad is cancelled, and these days cause most of the passenger delay cost. A schedule should therefore be tested against extreme days, instead of against an average day. The cost of this delay is expressed in euros for the passengers who are affected, but not for the loss of trust in public transport that follows from repeated cancellations. That means that the numbers understate the consequences for passengers. How the charging infrastructure explains the difference between the depots is discussed in section 6.4.9.

Conclusion on the benchmark schedule (SRQ3)

The benchmark schedules, which were operated in 2025, complete almost all trips at 5°C. At 0°C the completion rate drops to 99.4% in Apeldoorn and 96.4% in Lelystad, and at -5°C to 96.3% and 89.7%. Weighted over the scenarios, the expected completion rate is 99.61% and 98.26%, which corresponds to about 1,100 to 1,250 hours of passenger delay per year, with a societal cost of €36,700 and €40,700 and an expected cancellation cost for the operator of €18,300 and €22,500 per year. These are the costs within the scenarios that were considered. Extreme conditions have a large effect on the outcomes, even though they are infrequent. The benchmark schedule of Lelystad has 35 vehicle tasks fewer than that of Apeldoorn, but a similar passenger delay, which shows the benefit of opportunity charging at the hub.

The benchmark schedules are only one point of comparison. Whether a schedule that is planned differently would perform better, is assessed in the next section.

6.4. Sensitivity analysis

The outcomes that were presented in the sections above, especially the vehicle task residual distribution, all come together in the sensitivity analysis. In this analysis the benchmark schedule is compared with four alternative schedules, to answer how the EC input rate for fleet scheduling influences robustness and efficiency. As in the previous section, the unexplained variation in EC is included through a Monte Carlo simulation. In every run, the realised EC of each vehicle task is the expected EC from the model, multiplied by a random factor. This factor is drawn from the observed ratios between measured and predicted EC of vehicle tasks, as discussed in section 6.1.4. The simulation ran 1,000 runs of every scenario, for every schedule of every depot (90,000 simulations in total) and took 139.14 minutes to run.

6.4.1. Alternative schedules and correction of the benchmark

The names of the schedules refer to the EC input rate that was used to plan them, relative to the benchmark schedule (table 4.2). Schedule -20 is planned under the assumption that the EC is 20% lower than for the benchmark schedule. As the range of a bus is the usable battery energy divided by the EC per kilometre, this means that the bus is assumed to have a range that is about 25% longer. This makes the schedule in theory more efficient, because less charging is needed, but it contains less energy redundancy on cold days. Schedule $+10$, on the other hand, assumes a 10% higher EC, which is a range that is about 9% shorter, and should therefore be more robust but less efficient.

One important note has to be made about the planned schedule cost of the benchmark schedule, before the alternative schedules are compared with it. For both depots, the benchmark schedule is less efficient than the schedule that was planned with the assumption that EC rates are 10% higher. In Lelystad, for example, the vehicle tasks of the benchmark schedule take a total of 168.6 hours per day, against 156.6 hours for the -10 schedule and 164.9 hours for the $+10$ schedule. This does not reflect the overall trend between the schedules, where a higher EC input rate leads to a less efficient schedule. The explanation is the way in which the alternative schedules were generated. Although the settings of the Hastus for the alternative schedules were chosen to reflect the settings of the benchmark schedule, the tool still managed to create schedules that are more efficient. The differences between the schedules are therefore not fully attributable to the EC input rate. Comparing the alternatives with the benchmark would give too positive outcomes for more energy redundancy, as part of the cost reduction is an efficiency gap that has nothing to do with the EC value.

That is why the planned schedule cost of the benchmark is not used for the comparison of the planned cost. Instead, it is replaced by a reference: the average of the planned schedule cost of the -10 and $+10$ schedules, $C_{\text{ref}}^{\text{sched}} = \frac{1}{2}\big(C_{-10}^{\text{sched}} + C_{+10}^{\text{sched}}\big)$. The net change for the operator of schedule p then becomes

$$\Delta C_p^{\text{op}} = \big(C_p^{\text{sched}} - C_{\text{ref}}^{\text{sched}}\big) + \big(\overline{C}_p^{\,\text{disturbance}} - \overline{C}_{p_0}^{\,\text{disturbance}}\big), \tag{6.2}$$

where p_0 is the benchmark schedule. All other indicators, such as the cancellations and the passenger delay, are still compared with the benchmark. Note that the correction assumes a linear trend between the -10 and $+10$ schedules, so it remains an approximation. A dedicated optimisation model in which the EC input is the only changing parameter would remove this uncertainty. For transparency, the outcomes without this correction are given in appendix C. The benchmark schedule is still an important point of reference, as it is the schedule that was operated in 2025, and its cancellations are not affected by the correction.

6.4.2. Overall comparison without driver cost

Figure 6.14 and table 6.12 show how the alternative schedules perform compared with the benchmark schedule. The benchmark schedule itself is not included, as its outcomes are zero by definition. Every group of bars represents one schedule. The light bar is the change in passenger delay cost. The middle bar shows the change for the operator, consisting of the planned schedule cost (solid) and the expected cancellation cost (hatched), with the black line as their sum. The dark bar is the overall net change, where the whiskers show the 5^{th} to 95^{th} percentile over the Monte Carlo runs. The value at the end of each bar is the mean over all runs. All values are percentages of the annual operator cost of the benchmark schedule (the planned schedule cost plus the expected cancellation cost, €3.76 million for

Apeldoorn and €0.81 million for Lelystad), so a bar below zero is a saving. Because both the changes and the base are annual numbers for the same scenarios, they can be compared between schedules and depots. Personnel is not included in this first comparison. While the percentages do not translate to all the cost that an operator makes to run the fleet, they do give an idea of the order of magnitude between the depots themselves. The costs are the annualised cost for the days that the scenarios are expected to occur. The outcomes under other circumstances are not considered, and the expected annual costs mentioned in this section are thus a limited view on the actual yearly cost associated with fleet scheduling. Nevertheless, the results give an opportunity to assess what the consequences of choices on EC rates during the planning phase could be for real-time operations.

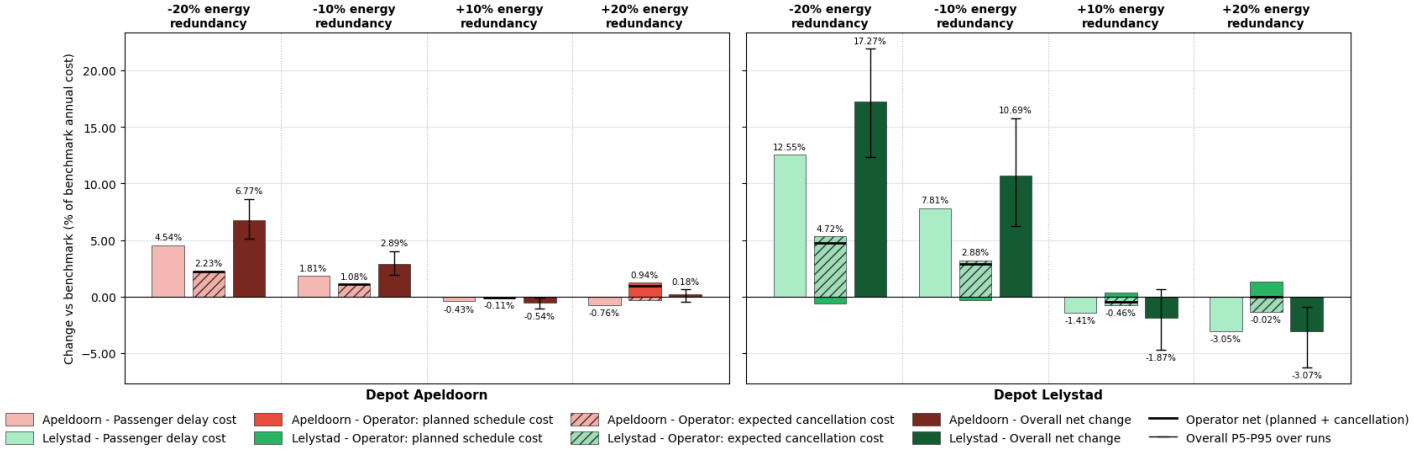


Figure 6.14: Net change in cost of the alternative schedules compared with the benchmark schedule, as a percentage of the annual operator cost of the benchmark schedule, without driver cost. The planned schedule cost is compared with the average of the -10 and $+10$ schedules

Table 6.12: Completion rate and net change in cost compared with the benchmark schedule, without driver cost. All values are percentages: the changes are percentages of the annual operator cost of the benchmark, and the planned schedule cost is compared with the reference described above

	Compl.	Pass.	Operator		Overall
Schedule	**rate (%)**	**delay**	**canc.**	**planned**	**change**
Apeldoorn					
-20	97.03	$+4.54$	$+2.74$	$+0.01$	$+7.29$
-10	98.30	$+1.81$	$+1.34$	$+0.01$	$+3.16$
Benchmark	99.61	–	–	–	–
$+10$	99.60	-0.43	-0.10	-0.01	-0.54
$+20$	99.85	-0.76	-0.34	$+1.23$	$+0.13$
Lelystad					
-20	93.76	$+12.46$	$+7.14$	-0.61	$+18.99$
-10	95.53	$+7.76$	$+4.31$	-0.33	$+11.74$
Benchmark	98.26	–	–	–	–
$+10$	98.89	-1.40	-1.05	$+0.33$	-2.11
$+20$	99.37	-3.03	-1.79	$+1.31$	-3.51

The figure and table show a clear picture. For both depots, planning with less energy redundancy leads to a higher overall cost: the overall net change is positive for the -10 and -20 schedules, in particular because of the passenger delay. The effect for more energy redundancy is much smaller, and not always an improvement. In Apeldoorn, only the $+10$ schedule leads to a (small) decrease in overall cost (-0.54%), as the $+20$ schedule, in spite of fewer cancellations, is slightly more expensive overall ($+0.13\%$) due to the extra vehicle task that it needs. In Lelystad, both the $+10$ and $+20$ schedules lead to a decrease in overall cost (-2.11% and -3.51%), as the lower passenger and operator cancellation cost outweigh the higher planned cost. The difference between the depots is substantial in terms of relative change in cost. This is an effect of the difference in charging infrastructure, which is explained

in detail in the following sections. Note that all numbers are for an average year of scenarios. They should be seen as an indication of the order of magnitude, rather than as a prediction.

Now it is time to dive deeper into how these metrics were calculated. First, robustness is thoroughly explained, followed by an analysis of how efficiency is affected under different EC rates. Last, the cost for the operator is assessed in more detail, including an estimate of the effects on driver cost.

6.4.3. Robustness assessment

For passengers, a higher reliability is always preferred, and for an operator it is the main indicator for the robustness of a schedule. In this section robustness is expressed in three ways: the average reliability, which is the scenario-weighted share of service trips that is completed, the worst-case reliability, which is the lowest reliability over all scenarios within a simulation run, averaged over the runs, and the passenger delay that results from the cancelled trips, in hours and in societal cost. The latter two are different in nature. The worst-case reliability shows how bad it can get on the most extreme day, while the delay shows how often and how severely passengers are affected over the course of a year.

Table 6.13 shows how the robustness indicators fluctuate amongst the different schedules for Apeldoorn (marked in red) and Lelystad (green). This colour scheme is featured throughout the tables in the remainder of this chapter.

Table 6.13: Reliability and delay performance for depots Apeldoorn (red rows) and Lelystad (green rows) across schedules

	Reliability		Delay		
Schedule	**Average reliability**	**Worst-case reliability (mean run min.)**	**Passenger delay (h/year)**	**Passenger delay cost (€/year)**	**Passenger delay cost difference (€/year)**
-20	97.03%	92.88%	6,395.86	207,655.61	170,927.95
-10	98.30%	93.60%	3,226.68	104,761.17	68,033.52
Benchmark	99.61%	96.13%	1,131.22	36,727.65	0.00
+10	99.60%	95.76%	629.69	20,444.19	-16,283.46
+20	99.85%	97.70%	250.37	8,128.68	-28,598.97
-20	93.76%	86.00%	4,368.25	141,824.77	101,141.03
-10	95.53%	84.44%	3,191.76	103,627.59	62,943.85
Benchmark	98.26%	89.37%	1,253.07	40,683.74	0.00
+10	98.89%	89.78%	903.20	29,324.32	-11,359.42
+20	99.37%	91.99%	495.34	16,082.23	-24,601.51

Both depots show a trend where the completion rate decreases for schedules that were planned with lower EC rates, and the differences get larger the more the EC input rate is lowered during planning (table 6.13 and table 6.12). In Apeldoorn the average completion rate drops from 99.61% for the benchmark to 98.30% for the -10 schedule and 97.03% for the -20 schedule. In Lelystad it drops from 98.26% to 95.53% and 93.76%. The effect is asymmetric: raising the EC input rate does not improve the completion rate of Apeldoorn (+10: 99.60%) or improves it only modestly (+20: 99.85%), while lowering the input rate has a large effect. This can be explained by the fact that it is hard to improve the completion rate once it is near 100%. In Lelystad, which has more room for improvement, the +10 and +20 schedules do increase it to 98.89% and 99.37%.

The worst-case reliability shows what a passenger can expect on the worst day. For the benchmark schedule, it is 96.13% for Apeldoorn and 89.37% for Lelystad. In other words, in the most extreme scenario about one in every ten service trips in Lelystad does not run. For the -20 schedule this rises to about one in seven (86.00%). The worst-case reliability does not decrease strictly with the input rate in Lelystad (-10 scores lower than -20), as every schedule is the outcome of a separate optimisation run, and differences in the structure of the vehicle tasks can influence how vulnerable they are on the most extreme days.

The passenger delay is already expected to occur under the benchmark schedule, at about 1,131 hours per year in Apeldoorn and 1,253 hours per year in Lelystad, equal to a societal cost of €36,728 and €40,684 respectively. For both depots the societal cost increases strongly for the -10 and -20 schedules (up to +€170,928 per year for Apeldoorn and +€101,141 for Lelystad, compared to the benchmark),

whereas the +10 and +20 schedules save €11,359 to €28,599 per year. As already seen for the benchmark schedules, the passengers of Lelystad and Apeldoorn experience a similar delay, in spite of the difference in size of the schedule. This is the benefit for passengers of having opportunity charging located at the hub of the network.

Overall, a lower EC input rate in the planning stage leads to a clear loss in robustness, while a higher input rate has diminishing returns. The question is what the robustness costs in terms of efficiency, which is the subject of the next section.

6.4.4. Efficiency assessment

Efficiency describes how much of the resources that are put into a schedule is used for the service that is offered. As the service trips are the same in all schedules, the service time and service distance are identical, and the differences are in everything that comes on top of that: the dead-heading trips, the charging events and the waiting time between trips. Efficiency is therefore expressed in three ways. The time efficiency is the share of service time in the driving time (service plus dead-heading) and in the total time of the vehicle tasks, which also includes charging and waiting. The distance efficiency is the share of service distance in the total driving distance. Besides that, the charging pattern is shown through the average charging duration, the number of charging requests per year and the number of requests that find all chargers occupied. The outcomes are listed in table 6.14, together with the operator cost in table 6.15.

The influence of the fast chargers at the hub is again clear in Apeldoorn. It is easy to find a gap in a vehicle task to charge in between trips, so there is little variation between the schedules. The distance efficiency hardly moves (93.4% for all schedules), as the dead-heading distance is about the same, while the total vehicle task time still changes a bit: from 689 hours per day for the -20 schedule to 705 hours for the benchmark. The only noticeable step is the +20 schedule, which needs one extra vehicle task. This adds around 10 hours of vehicle task time per day compared to the +10 schedule and lowers the share of service time in the total task time from 75.0% to 73.9%.

For Lelystad the results are much more telling. Every trip back to the depot is about six kilometres, so an extra charging moment costs six kilometres of dead-heading, plus the time that is needed to drive these kilometres. The distance efficiency decreases from 90.1% for the -20 schedule to 85.9% for the +20 schedule, and the total vehicle task time increases from 155.0 to 170.7 hours per day (+10%) while the service time stays the same. The share of service time in the total task time drops from 76.2% to 69.2%. The benchmark schedule is again an outlier here: it is less efficient than the +10 schedule, which confirms the earlier remark about the benchmark.

The charging pattern shows a few more things. In Apeldoorn the benchmark schedule has many short charging events (about 7,271 requests per year at an average of 46 minutes), whereas the alternative schedules have fewer and longer events. More importantly, there is a problem that is typical for opportunity charging at a limited number of chargers. When buses have to charge earlier than planned due to a higher EC, there are cases in which all chargers are occupied (the last column of table 6.14: the number of times per year that a bus has to go to a charger earlier than planned, but finds no charger available). In the -20 schedule this is expected to happen 118 times per year, and for the -10 schedule 49 times, against 17 times for the benchmark. In absolute terms this is not much, but it shows that there are limits to opportunity charging. Such situations can lead to considerable inconvenience for drivers, as a bus has to wait for a charger, or has to be replaced. This does not occur in Lelystad, where chargers are located at the depot, and where there are 10 chargers for 10 vehicle tasks.

To conclude, the schedules with a higher EC input rate are in general less efficient, as expected, but the differences are small for Apeldoorn and larger for Lelystad. These numbers do not say everything. The efficiency of a vehicle schedule is closely related to the efficiency of the driver schedules, as a driver can only work on a bus when it is available. Extra dead-heading, longer charging and extra vehicle tasks can result in more complex or less efficient rosters. This is discussed in the next section, where the effects on the cost for the operator are assessed.

Table 6.14: Efficiency performance for depots Apeldoorn (red rows) and Lelystad (green rows) across schedules

	Time						Distance			Charging		
Schedule	**Vehicle tasks**	**Service time (h/day)**	**Driving time (h/day)**	**Service / driving (%)**	**Total task time (h/day)**	**Service / total (%)**	**Service distance (km/day)**	**Total driving distance (km/day)**	**Service distance efficiency (%)**	**Avg. charging duration (min)**	**Annual charging requests**	**Requests w/ all chargers occupied**
-20	45.00	518.42	556.27	93.20	689.03	75.24	15,818.13	16,930.77	93.43	70.46	5,350.08	117.74
-10	45.00	518.42	556.25	93.20	689.93	75.14	15,818.13	16,928.45	93.44	63.96	5,290.95	49.45
Benchmark	45.00	518.42	558.93	92.75	705.47	73.49	15,815.31	16,929.08	93.42	45.96	7,270.53	17.03
+10	45.00	518.42	556.63	93.13	691.18	75.00	15,818.13	16,923.12	93.47	55.02	5,984.34	7.53
+20	46.00	518.42	557.83	92.93	701.10	73.94	15,818.13	16,933.57	93.41	50.84	7,367.39	4.03
-20	10.00	118.08	125.87	93.82	155.00	76.18	2,881.74	3,197.84	90.12	97.86	868.36	0.00
-10	10.00	118.08	126.42	93.41	156.57	75.42	2,881.74	3,220.27	89.49	96.75	820.71	0.00
Benchmark	10.00	118.08	128.97	91.56	168.60	70.04	2,881.21	3,327.64	86.58	90.36	1,054.22	0.00
+10	10.00	118.08	127.67	92.49	164.85	71.63	2,881.74	3,274.24	88.01	107.82	805.30	0.00
+20	10.00	118.08	129.52	91.17	170.70	69.18	2,881.74	3,353.49	85.93	93.94	1,109.45	0.00

Table 6.15: Operator cost performance for depots Apeldoorn (red rows) and Lelystad (green rows) across schedules, without driver cost. The planned cost is compared with the average of the -10 and $+10$ schedules, the cancellation cost with the benchmark schedule

	Planned annual cost				Expected annual cancellation cost					Total difference
Schedule	**Fixed cost (€)**	**Variable cost (€)**	**Total schedule cost (€)**	**Planned cost difference with reference (€)**	**Cancellation effect variable cost (€)**	**Missed income cost (€)**	**Fine cost (€)**	**Cancellation cost (€)**	**Cancellation cost difference (€)**	**Total operator cost difference (€)**
-20	2,053,125.00	1,693,076.90	3,746,201.90	498.40	-13,509.40	89,439.86	45,301.99	121,232.45	102,974.96	103,473.36
-10	2,053,125.00	1,692,845.30	3,745,970.30	266.80	-6,620.38	49,408.79	25,956.87	68,745.28	50,487.80	50,754.60
Benchmark	2,053,125.00	1,692,908.10	3,746,033.10	–	-1,833.98	14,208.55	5,882.92	18,257.48	0.00	–
+10	2,053,125.00	1,692,311.70	3,745,436.70	-266.80	-1,308.04	9,660.94	6,076.23	14,429.13	-3,828.36	-4,095.16
+20	2,098,750.00	1,693,356.90	3,792,106.90	46,403.40	-511.93	3,701.64	2,293.76	5,483.47	-12,774.02	33,629.38
-20	456,250.00	319,784.30	776,034.30	-4,941.25	-2,747.06	41,636.44	41,561.06	80,450.44	57,915.95	52,974.70
-10	456,250.00	322,026.80	778,276.80	-2,698.75	-2,003.74	29,738.44	29,789.83	57,524.53	34,990.04	32,291.29
Benchmark	456,250.00	332,763.90	789,013.90	–	-806.16	11,725.42	11,615.23	22,534.49	0.00	–
+10	456,250.00	327,424.30	783,674.30	2,698.75	-486.29	7,100.76	7,431.13	14,045.60	-8,488.88	-5,790.13
+20	456,250.00	335,349.30	791,599.30	10,623.75	-303.24	4,155.45	4,169.23	8,021.44	-14,513.05	-3,889.30

6.4.5. Cost assessment

The efficiency of a schedule translates into cost in two ways. The planned schedule cost consists of the fixed cost per bus and the variable cost per kilometre, which are affected by the number of vehicles and the dead-heading kilometres. In addition, there is the expected cost of cancelled trips, which consists of the income that is missed for the service hours that are not driven, the fines, and the variable cost that is saved by not driving the trips. Table 6.15 shows the composition of these costs for both depots.

The first thing that stands out in the table is that the planned schedule costs hardly change between the schedules. In Apeldoorn the difference with the reference is less than €500 per year for all schedules except the +20 schedule, which is about 0.01% of the annual operator cost. The exception is the extra vehicle task of the +20 schedule, which costs €45,625 in fixed cost per year, and is therefore the only schedule that changes the planned cost considerably (+€46,403). In Lelystad the planned cost differs more, from €4,941 less for the -20 schedule to €10,624 more for the +20 schedule, mostly via the variable cost for the dead-heading distance. This confirms that the largest gain is not in the planned cost, but in preventing the loss of income for cancelled trips.

The cancellation cost is mostly determined by the missed income and the fines. In Apeldoorn, the missed income of the benchmark schedule is €14,209 per year against €5,883 for fines, and in Lelystad €11,725 against €11,615. The share of the fines is higher in Lelystad because the average trip is shorter, and a trip therefore represents less income, whereas the fine is a fixed amount per trip. The cancellation cost increases quickly for lower input rates. In Apeldoorn it rises from €18,257 for the benchmark to €68,745 for the -10 schedule and €121,232 for the -20 schedule. With the exception of the extra vehicle task of the +20 schedule, the total operator cost therefore follows the same pattern as the cancellation cost: the schedules with a lower EC input rate cost the operator €50,755 (-10) and €103,473 (-20) per year more in Apeldoorn, and €32,291 and €52,975 in Lelystad. In Apeldoorn the +10 schedule is the only one that leads to a (small) saving of €4,095 for the operator, as the extra vehicle task of the +20 schedule (+€46,403 planned cost) costs more than the roughly €41,000 per year of lower cancellation and passenger delay cost that it brings. In Lelystad both the +10 and +20 schedules save money for the operator (€5,790 and €3,889), as the lower cancellation cost outweighs the higher planned cost. These savings are small in absolute terms, so the case for more energy redundancy rests mostly on the benefit for the passengers.

The amount of the fine is chosen rather arbitrarily, and needed to be simplified to scale well with the size of the depot (see chapter 4). When fines increase, an operator would be more inclined to prevent cancellations, so the benefit of a more robust schedule would increase. The effect of this choice is tested in section 6.4.7.

The costs in this section are limited to the vehicle schedule. As already hinted in the previous section, the efficiency of a vehicle schedule is closely linked to the efficiency of the driver schedules, and the cost of drivers is the largest cost component for an operator. The next section therefore adds an illustration of the effect of these cost.

6.4.6. Including driver cost

Personnel is included in this section to illustrate the possible consequences of efficiency improvements, or losses, in the vehicle schedule. Generally, the cost for direct personnel accounts for almost two thirds of the total cost that a public transport operator makes: the wage component has a weight of 64.3% in the national contribution index (LBI) with which the payments in bus concessions are indexed for battery-electric buses (DOVA, 2022, 2025), and CROW-KpVV and Ecorys (2021) reports a share of 56% to 66% for personnel in bus, tram and metro concessions. A gain at fleet level can therefore easily be undone by a relatively small increase in driver cost, and more dead-heading, charging and waiting also makes it harder to build efficient driver schedules. Because no driver schedules are available for the alternative vehicle schedules, the effect is estimated from the hours of the vehicle tasks. One hour in the daily schedule is valued at €20,000 per year. This amount is derived from the annual cost of €60,000 to €65,000 per full-time driver in the CROW key figures (CROW-KpVV & Ecorys, 2021), taking the midpoint of €62,500. A schedule for a normal weekday is operated on five days a week, so one hour in the daily schedule corresponds to five hours of service per week. A driver is on average deployable on 3.5 days per week, and of an eight-hour working day about 5.5 hours are actually spent driving (EBS, internal data). One driver can therefore cover about 19.25 hours of service per week, so

one hour in the daily schedule requires about 0.26 drivers. This results in about €16,200 per year at the price level of the key figures, which is about €20,000 after adjusting to the price level of 2026.

All driving time (service and dead-heading) is valued at this rate. The difference between the schedules is thus also found in the dead-heading hours. The idle hours within a vehicle task (charging and waiting) are valued at 25% of the rate for driving hours, so at €5,000 per year per hour in the daily schedule. This is an assumption: it reflects that more idle time makes it harder to build efficient driver schedules, without all of that time being paid work. It does not take into account the actual effects on the crew schedules, which would require a crew scheduling step. The influence of this assumption is tested in section 6.4.7.

Figure 6.15 adds the estimated driver cost to the planned schedule cost, and is the counterpart of figure 6.14. The reference for the planned cost is the same as before, so it now includes the driver cost of the -10 and $+10$ schedules. To keep the outcomes comparable with those without driver cost, the percentages are still expressed relative to the annual operator cost of the benchmark schedule without personnel. These percentages would be much lower if personnel and overhead were part of the base. The figure also shows the range of the outcomes for different cost assumptions, which is discussed in section 6.4.7. The outcomes are listed in table 6.16.

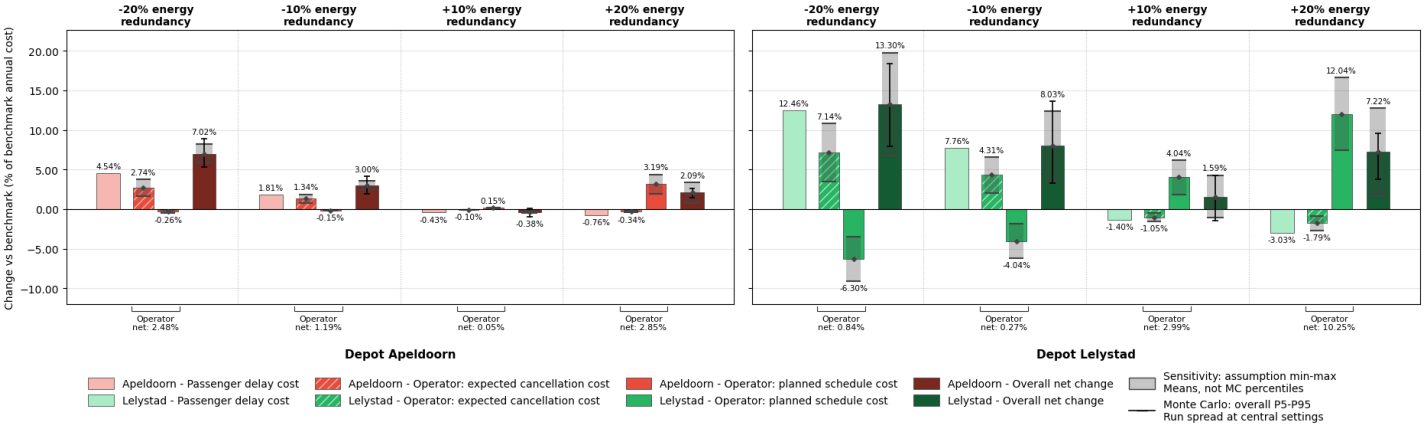


Figure 6.15: Net change in cost compared with the benchmark schedule with the estimated driver cost included in the planned schedule cost, as a percentage of the annual operator cost of the benchmark schedule without driver cost. The grey bars show the range of the overall net change for different fines and values of idle hours

Table 6.16: Net change in cost compared with the benchmark schedule, including driver cost. All values are percentages of the annual operator cost of the benchmark without driver cost; the planned schedule cost is compared with the reference described in section 6.4.1. The cancellation cost and the passenger delay cost do not change by including driver cost, so the latter is only given in table 6.12 (the overall change includes it). The range is that of the overall change for fines of €0 to €40 per cancelled trip and idle hours valued at 0% to 50% of the rate for driving hours

	Operator		Overall	Overall
Schedule	**canc.**	**planned**	**change**	**range**
Apeldoorn				
-20	$+2.74$	-0.26	$+7.02$	$[5.79, 8.24]$
-10	$+1.34$	-0.15	$+3.00$	$[2.41, 3.59]$
$+10$	-0.10	$+0.15$	-0.38	$[-0.44, -0.32]$
$+20$	-0.34	$+3.19$	$+2.09$	$[0.78, 3.40]$
Lelystad				
-20	$+7.14$	-6.30	$+13.30$	$[6.82, 19.78]$
-10	$+4.31$	-4.04	$+8.03$	$[3.62, 12.43]$
$+10$	-1.05	$+4.04$	$+1.59$	$[-1.09, 4.28]$
$+20$	-1.79	$+12.04$	$+7.22$	$[1.67, 12.77]$

With the driver cost included, the overall net change in Apeldoorn is $+7.0\%$ and $+3.0\%$ for the -20 and -10 schedules, -0.4% for the +10 schedule and $+2.1\%$ for the +20 schedule. In Lelystad it is $+13.3\%$,

$+8.0\%$, $+1.6\%$ and $+7.2\%$. The schedules with less energy redundancy remain costly, even though their planned schedule cost is lower (up to -6.3% in Lelystad), because the extra cancellation and passenger cost is larger. For the operator alone they are almost cost-neutral in Lelystad ($+0.8\%$ and $+0.3\%$), so the cost is mainly carried by the passengers. For more energy redundancy the picture changes. In Lelystad the savings on cancellations and passenger delay of the +10 schedule are outweighed by an increase in planned schedule cost of 4.0%, so it no longer pays off, and the +20 schedule becomes much more expensive ($+7.2\%$). In Apeldoorn the +10 schedule remains about cost-neutral, but the +20 schedule no longer pays off either. The extra driving and idle hours of a more robust schedule thus cost more than what is saved on cancellations, except for the +10 schedule in Apeldoorn. This also shows why an operator can be hesitant to improve robustness without knowing the effects on the driver schedules.

The cancellation cost and the delays of passengers are calculated without taking further consequences into account. A cancelled bus could also mean a missed connection to a train for a passenger, or a driver who is left without a bus, because it is at a charger elsewhere. In the long term, this can lead to passengers avoiding public transport, and to more stress among drivers. These considerations plead for a careful trade-off that leans towards more robustness, and saving cost by accepting additional cancellations should never be the starting point.

6.4.7. Sensitivity to the cost assumptions

The outcomes depend on two simplified assumptions: the fine per cancelled trip and the driver cost for idle hours. To see how much influence they have, both are varied without a new simulation. The fine ranges from €0 to €40 per cancelled trip, and idle hours are valued at 0% to 50% of the rate for driving hours. Figure 6.15 and table 6.16 show the range of the overall net change over the 15 combinations. This makes the findings of the previous sections more precise. The finding that less energy redundancy is costly holds in all combinations: the net change of the -20 and -10 schedules is positive, between $+2.4\%$ and $+8.2\%$ in Apeldoorn and between $+3.6\%$ and $+19.8\%$ in Lelystad. The finding that more energy redundancy pays off does not. The +20 schedule is more expensive than the benchmark in both depots in all combinations. The +10 schedule stays slightly below zero in Apeldoorn (-0.4% to -0.3%), but in Lelystad it ranges from -1.1% to $+4.3\%$, so there it only pays off for part of the assumptions. The two assumptions work in opposite directions. A higher fine makes cancellations more expensive, which benefits a more robust schedule, whereas a higher value for idle hours mainly penalises the schedules with more charging and waiting time. The long-term effects of unreliable service, such as passengers who stop using public transport, are not part of these numbers. Including them would favour a more robust schedule.

6.4.8. Seasonal schedules

Because the effect of temperature is so dominant, a distinction between a winter and a summer schedule could be considered, so that the energy redundancy is only applied in the period that it is needed. A quick check with the existing simulation results shows what this would require. Even at 5°C, the -20 schedule leads to about 23 to 24 cancelled trips per day in both depots, where this is about 0.2 trips in Apeldoorn and 1.4 trips in Lelystad for the benchmark schedule. The -10 schedule still has 8 cancelled trips in Apeldoorn and 10 in Lelystad. In almost every run of these schedules, trips are cancelled at 5°C, so a schedule with less redundancy cannot be used in the period in which such temperatures occur. At that temperature, the societal and operator cost of the -20 schedule is about €2,600 per day in Apeldoorn and €2,400 per day in Lelystad, whereas the estimated saving on planned schedule cost (including driver cost) is only about €50 and €250 per day respectively. A summer schedule could therefore only pay off on days that are considerably warmer than the scenarios that were simulated here, and the benefit is limited by the number of days on which this applies. The temperature at which a schedule with less redundancy no longer leads to cancellations was not simulated, and would require scenarios with higher temperatures. It is therefore a promising search direction for an operator rather than a conclusion. The drawbacks should be weighed as well: an extra schedule means extra work for the planners, and a change of the duties of the drivers twice a year, which drivers are generally not pleased with.

6.4.9. Reflection on the trade-off

The outcomes can be placed in the context of earlier work. The trade-off between schedule efficiency and robustness has been quantified before by Yap and Van Oort (2018), who compared two driver schedule types for a light rail disruption in The Hague. They weighed the additional scheduling cost of a more robust schedule (€300,000) against the reduction in passenger and operator disruption costs, and found that the total benefit (€500,000) exceeded the additional cost, whereas the benefit for the operator alone (€200,000) did not. This study extends that approach to vehicle scheduling under EC uncertainty, and evaluates multiple scenarios with a probability instead of a single disruption. The overall conclusion is similar: more robustness pays off for society. For the operator the picture depends on the assumptions. Without driver cost, the operator also benefits from the +10 schedule, as the cost of a cancelled trip (missed income and fines) is larger than the extra cost of dead-heading and charging. With driver cost, this advantage largely disappears, which is in line with the finding of Yap and Van Oort (2018) that the benefit for the operator alone is not clear-cut. The cost of energy redundancy is also much smaller than the 67% extra cost that Momen et al. (2025) find for designing charging infrastructure for the worst-case EC. That is not surprising: this study considers the choice of a planning value within an existing infrastructure, where extra redundancy can partly be absorbed by the schedule, whereas the design of infrastructure requires physical investments. It fits the result of Tang et al. (2019) that a buffer distance can prevent breakdowns while maintaining cost efficiency. Finally, the approach is comparable to the DES with Monte Carlo simulation of De Bruin et al. (2023), who find a trade-off between driving time reliability and cost. The difference is that this study uses an EC model that is based on real data and includes the effects of temperature, and that the consequences are valued for passengers and the operator.

The difference between the depots is large, and it is mainly an effect of the charging infrastructure. Apeldoorn has fast chargers at the hub, and Lelystad has slower chargers at the depot, six kilometres from the hub. Two aspects are confounded in this comparison: the location of the chargers (hub or depot) and their power (230 or 100 kW), and with two depots they cannot be separated. It is plausible, but not tested, that more chargers, faster chargers or a depot closer to the hub reduce the cost of redundancy, and that a distant depot increases it. A bus that needs energy earlier than planned can charge during a layover at the hub, instead of driving to the depot, and after a disturbance it is back in service sooner. Opportunity charging at the hub also has limits, as the capacity of the chargers is finite: with a lower EC input rate, the number of times that all chargers are occupied increased strongly. Conclusions about other depots should therefore be seen as hypotheses, which require more case studies. In the future, a larger battery capacity could make the problem less relevant, as a bus that can drive a vehicle task on a single charge, even on a cold day, does not depend on charging during the day.

Several limitations should be kept in mind. First, the alternative schedules were generated with Hastus, which is a black box in this research. Although it is a real-world example of an optimisation technique that is used by an actual operator, the process cannot be explained in detail, and the differences between the schedules are not fully attributable to the EC input rate. The correction of the benchmark is an approximation for this. Designing a dedicated optimisation model, in which the input rate is the only parameter that changes, would remove this uncertainty. Second, the simulation uses simple decision rules and no real-time recovery measures, and a single random factor per vehicle task. This keeps the runtime acceptable (139 minutes for 90,000 simulations), but it does not capture the variation between trips within a task or correlations with the temperature. Other studies deal with the unexplained variance in different ways. Bie et al. (2021) use it as a probability distribution within the optimisation, and Momen et al. (2025) construct an ambiguity set around the empirical distribution to deal with sparse data. Third, the financial numbers depend on assumptions, such as the fine per cancelled trip, which differs strongly between Dutch concessions, and the driver cost, which was added as an illustration. Although the societal cost are included, a systematic drop in reliability can cause passengers to avoid public transport in the future, which results in lower ticket revenue. Cancelled trips can also lead to missed connections, or to drivers who are left without a bus. These effects are not included. Finally, the level of service, in the form of frequencies, is left out of scope. A robust schedule that needs more buses or kilometres could, in a setting with a fixed budget, lead to less service elsewhere. The outcomes should therefore be seen as one input in the discussion about robustness, not as the answer to it.

Conclusion on the influence of the EC input rate (SRQ4)

The trade-off between robustness and efficiency is asymmetric. Planning with less energy redundancy (a lower EC input rate) reduces the completion rate sharply and raises the cost for passengers and the operator, whereas planning with more redundancy has diminishing returns. In Apeldoorn the overall cost of the -10 and -20 schedules is 3.2% and 7.3% higher than that of the benchmark, in Lelystad 11.7% and 19.0%. The +10 schedule leads to savings without driver cost (-0.5% and -2.1%), but the benefits become small or disappear when driver cost is included (-0.4% in Apeldoorn and $+1.6\%$ in Lelystad), and the +20 schedule does not pay off in either depot. Charging infrastructure explains most of the difference between the depots: with fast chargers at the hub, the cost of redundancy is small. The outcomes depend on the cost assumptions, but the finding that less energy redundancy is costly holds for all combinations that were tested.

The answers to the sub-questions are combined in the next chapter, where the main research question is answered and the general findings are formulated.

7

Conclusion

This study aimed to quantify how the energy consumption (EC) value that is assumed in the scheduling of electric buses shapes the trade-off between robustness and efficiency, for public transport facing EC uncertainty. The sub-research questions (SRQs) were answered, and compared with the literature, in the sections of the previous chapter. This chapter combines these answers into an answer to the main research question, and then looks beyond the case study. It names the scientific contributions, formulates the general findings that are expected to hold for other concessions and operators, and closes with recommendations for practice and for further research.

7.1. Answers to the research questions

External factors (SRQ1)

What external factors influence energy consumption of battery electric buses in a Dutch PT network and what is the impact on energy consumption for each factor?

Temperature is by far the most important factor for the EC of an electric bus. The relationship is U-shaped with an optimum of about 19°C, and at the coldest temperatures that occur regularly the EC is 35% to 48% higher. This means that a bus has roughly a quarter to a third less range than at the optimum. The state of charge (SoC) at the start of a trip and the number of passengers follow at a large distance. The regression model that was developed on a year of real trips explains 76% of the variance on vehicle task level, but only 27% (12m bus) and 47% (13m bus) on trip level, as the integer SoC in the data causes large errors for short trips. On vehicle task level, the model is accurate enough (with a mean absolute error of about 5% of the average EC) to assess schedules. The remaining unexplained variance of roughly 10% is as large as the redundancy that is studied, so it has to be part of the simulation.

Scenarios (SRQ2)

What scenarios can be defined to test the robustness of a schedule for a battery electric bus fleet in a Dutch context?

Temperature and the number of passengers are the two factors that define the scenarios. For this Dutch case study, three relevant temperatures (-5°C, 0°C and 5°C), which occur on about 67 days per year in the expected climate of 2027, are combined with three levels of passengers, which results in nine scenarios in total. Each scenario has a number of days that it is expected to occur in a year, which can be used to annualise results. Cold days are also busy days, which is the most harmful combination when trips are cancelled. The scenarios describe an average year, and do not cover warm days or a series of extreme winters.

Robustness of the benchmark schedule (SRQ3)

What is the estimated robustness of an actual schedule when values of external factors change according to the given scenarios?

The schedules that were operated in 2025 are robust down to about 5°C. At 0°C the completion rate is 99.4% in Apeldoorn and 96.4% in Lelystad, and at -5°C it drops to 96.3% and 89.7%. In Lelystad this

means that about one in ten service trips is cancelled on the coldest days, often in the evening when the next bus is far away. Over the scenarios this adds up to about 1,100 to 1,250 hours of expected passenger delay per year (a societal cost of about €37,000 to €41,000), and an expected cancellation cost for the operator of about €18,000 to €23,000 per year. Cancelled trips make public transport less reliable, which means that it is valued less by passengers. These numbers are the trips that are at risk when no recovery measures such as a bus swap are available. Although temperatures of -5°C occur on less than 2% of the days, they cause a substantial part of the cost: about 63% of the passenger delay cost in Apeldoorn and 38% in Lelystad. Apeldoorn, with fast chargers at the hub, has a similar passenger delay as Lelystad with 35 more vehicle tasks.

Influence of the EC input rate (SRQ4)
How does a change in the energy consumption input rate for fleet scheduling influence the robustness and efficiency of the given output?
The effect of the EC value that is assumed in planning is asymmetric. Planning with less energy redundancy (a lower value) reduces the completion rate sharply and raises the overall cost for passengers and operator by 3.2% to 7.3% in Apeldoorn and by 11.7% to 19.0% in Lelystad, relative to the annual operator cost of the benchmark. Planning with more redundancy has diminishing returns. A 10% higher value saves 0.5% (Apeldoorn) and 2.1% (Lelystad) without driver cost. For the operator alone this is only about €4,100 and €5,800 per year, so the benefit lies mostly with the passengers. A 20% higher value only pays off in Lelystad. In Apeldoorn the extra vehicle task that it needs costs about €46,000 per year, more than the roughly €41,000 of cancellation and passenger delay cost that it saves. The percentages change noticeably when the estimated driver cost is included, as more charging, deadheading and waiting make the driver duties less efficient. The extra cost of less redundancy then becomes 3.0% to 7.0% in Apeldoorn and 8.0% to 13.3% in Lelystad, and the benefit of a 10% higher value becomes -0.4% in Apeldoorn and disappears in Lelystad ($+1.6\%$), where a 20% higher value is now more expensive than the benchmark as well ($+7.2\%$). With a 20% lower input value, the chargers in Apeldoorn are fully occupied about 118 times per year when a bus needs to charge earlier than planned, which shows that opportunity charging at the hub has its limits.

Main research question
How does the energy consumption value that is assumed in the scheduling of electric buses shape the trade-off between robustness and efficiency, for public transport facing energy consumption uncertainty?
The EC value that is assumed in the scheduling shapes the trade-off asymmetrically. Too little energy redundancy leads to cancelled trips on cold days, of which the cost for passengers and the operator quickly exceeds the savings on the schedule. Too much redundancy results in a rapidly diminishing return on the completion rate and in extra planned schedule cost, that is not earned back. In other words, cutting robustness to save costs can end up costing more. Whether more redundancy than the current planning value pays off depends on the effect on driver schedules as well, while the finding that less redundancy is costly holds for all cost assumptions that were tested. The results also show that the exact numbers on this trade-off are dependent on the charging infrastructure, effect on crew scheduling efficiency and the height of fines for cancelled trips. In this case study, a 10% higher value is the best of the options that were tested when driver cost is left out, but it is not necessarily the optimum, and the best level depends on the depot. As the framework can be applied to other concessions with the data that operators already collect, it provides a starting point for the discussion on the desired level of robustness.

7.2. Scientific contributions

The outcomes follow from three scientific contributions, each of which addresses a gap in the literature (chapter 2). First, a data-driven EC model was developed on a full year of real trips and validated on vehicle task level, whereas the reviewed studies report the accuracy per trip or at an even finer resolution. It shows that a model that explains only part of the variance per trip can still be accurate enough to assess a schedule, as the errors of individual trips partly cancel each other out. It also shows that this can be achieved with data that operators already collect, in contrast to the high-frequency data that the most accurate models use. Second, a scenario-based simulation framework was introduced that combines this model with scenarios for temperature and passengers, and with the unexplained

variance (Monte Carlo), to test the robustness of a whole electric bus schedule under rare but impactful conditions. Third, the trade-off between robustness and efficiency was quantified for the EC value that is used in vehicle scheduling, for passengers and the operator, which was not found in the literature. This includes a correction that separates the effect of the EC value from differences in how the schedules were generated, and a test of how the conclusions depend on the cost assumptions.

7.3. General findings

Table 7.1: General findings derived from the case study

#	Conclusion	Case study outcome	Mechanism and conditions
1	The climate of a concession determines when a schedule is at risk.	Temperature explains most of the variance in EC, and cold days are also the busiest days.	Applies to regions with a similar climate. In warmer regions, cooling at high temperatures may become the risk, which the data could not confirm.
2	Too little energy redundancy costs more than it saves, and the effect is asymmetric.	Lower EC value: $+1.8\%$ to $+12.5\%$ passenger delay cost. Higher EC value: only -0.4% to -3.0%.	Cancellations carry cost for passengers and operator, while a completion rate near 100% leaves little to gain. Applies where cold weather raises the EC considerably.
3	Fast charging at the hub dampens the effect of energy redundancy, both in planning and in recovery during operations.	Apeldoorn has a similar passenger delay as Lelystad with 35 more vehicle tasks, and almost no extra planned cost for extra redundancy.	In planning, energy is restored quickly in a layover that is part of the service, without dead-heading. In operations, a bus that runs low can be recharged at the hub within the service and is back in service sooner, which limits the cancellations and shortens the recovery. Confounded by charger power, limited by the capacity of the chargers, and only two depots.
4	The value of energy redundancy depends on the ratio between cancellation cost and the efficiency loss of redundancy.	The +10 schedule saves cost without driver cost (-0.5% in Apeldoorn and -2.1% in Lelystad). The fine and the value of idle time shift the outcome.	Redundancy pays off where fines and missed income are high and extra charging or vehicles are cheap.
5	Crew scheduling can decide whether robust fleet schedules pay off.	With driver cost, the net change of the +10 schedule is -0.4% in Apeldoorn and $+1.6\%$ in Lelystad.	More charging, dead-heading and idle time make driver schedules less efficient. Operators may be hesitant to improve robustness without knowing the effect on crew schedules.
6	Rare extreme days dominate the cost of disturbances, so schedules should be tested against extremes.	Days of -5°C occur on less than 2% of the days, but cause 63% (Apeldoorn) and 38% (Lelystad) of the passenger delay cost.	Cancelled trips on busy, cold days affect many passengers at once. Applies to every concession where one condition drives the energy consumption, but the extremes have to be weighted with their probability.

The outcomes are those of two depots, so the question is what they mean for other concessions and operators. The general findings in table 7.1 combine the outcomes of the case study with knowledge about the mechanisms behind them and an interpretation of what this would mean elsewhere. The conditions under which each finding is expected to hold are stated, so that the reader can judge whether it applies to another concession.

Findings 2 and 4 together give an operator a way to judge whether the case study applies to its own situation. Redundancy pays off when the cost of a cancelled trip, which consists of the passenger delay, the missed income and the fine, is high compared with the cost of the extra charging, dead-heading and vehicles that it requires. The asymmetry follows from this. When the completion rate is already close to 100%, extra redundancy can only prevent a few cancellations, whereas too little redundancy causes many cancellations on days that are both cold and busy.

Findings 3 and 5 are conditions rather than results, as they determine how large the cost of redundancy is. Fast charging at the hub lowers it, as energy can be added within the service, whereas a distant depot turns every extra charge into a long absence of the bus. Crew scheduling can even reverse the outcome, as the case study shows for Lelystad, where the benefit of a 10% higher value disappears when the effect on the driver duties is included. Operators therefore have to look at the charging infrastructure and at the driver schedules before they can judge whether a more robust vehicle schedule

pays off.

Two further insights do not follow from the comparison of the schedules, but are worth mentioning. First, a larger battery capacity could make the problem less relevant in the future. A bus that can drive a complete vehicle task on a single charge, even on a cold day, does not depend on charging during the day, and the need for energy redundancy in the planning would decline accordingly. Second, a schedule with less redundancy during warmer months could save cost during that period, without harming robustness. It is not studied what the temperature threshold should be and for what period it should hold. This makes a seasonal schedule a possible search direction for an operator, with the drawbacks of extra planning work and a change of the driver duties twice a year.

7.4. Recommendations for practice

For EBS Public Transportation (EBS) and other operators, the outcomes lead to three recommendations. First, study the local conditions in which EC rates can jump and the probability of such events, together with what decides the cost of redundancy at the depot: the location and capacity of the chargers, the cost of a cancelled trip (passenger delay, missed income and fines) and the effect on the driver schedules. As the value of redundancy depends on this local context, the desired redundancy has to be formulated for a specific depot, instead of copied from another one. Second, use the framework as a stress test before a concession is operated or a new schedule is implemented. It shows what a schedule does on cold days and what energy redundancy costs, using data that operators already collect, and it allows planners to explain to internal stakeholders how choices on robustness are made. Third, improve the data on EC, in particular the resolution of the SoC, as this reduces the unexplained variance, and with it the redundancy that is needed. The financial numbers should be used as one input in the discussion about robustness. The long-term effects of unreliable service, such as passengers who stop using public transport, are not included and plead for a choice that leans towards more robustness.

7.5. Limitations and recommendations for further research

The limitations were discussed in chapter 6, and lead to the following recommendations for further research. A dedicated optimisation model, in which the EC value is the only parameter that changes, would make the differences between schedules fully attributable to that value, and remove the need to correct the benchmark. Applying the framework to more depots and regions, with different charging infrastructure, climates and trip lengths, would test the general findings, and would make it possible to separate the effect of charger location from the effect of charger power. Scenarios that cover sequences of years, for example several mild years followed by an extreme winter, would show how sensitive the outcomes are to the variation between years, and scenarios with higher temperatures would show when a seasonal schedule starts to pay off. A richer EC model that includes the road gradient, the number of stops, an occupancy rate and the speed dynamics (appendix B) could reduce the unexplained variance. A more realistic simulation, with real-time recovery measures such as bus swaps and variation between trips within a vehicle task, would make the outcomes less sensitive to the simplifications. Finally, including a crew scheduling step and the effects on the level of service and on long-term demand would complete the picture of the trade-off.

References

Abdelaty, H., Al-Obaidi, A., Mohamed, M., & Farag, H. E. Z. (2021). Machine learning prediction models for battery-electric bus energy consumption in transit. *Transportation Research Part D: Transport and Environment*, *96*, 102868. https://doi.org/10.1016/j.trd.2021.102868

Abdelaty, H., & Mohamed, M. (2021). A prediction model for battery electric bus energy consumption in transit. *Energies*, *14*, 2824. https://doi.org/10.3390/en14102824

Alamatsaz, K., Hussain, S., Lai, C., & Eicker, U. (2022). Electric bus scheduling and timetabling, fast charging infrastructure planning, and their impact on the grid: A review. *Energies*, *15*, 7919. https://doi.org/10.3390/en15217919

Aparício, J. T., Arsénio, E., & Henriques, R. (2022). Assessing robustness in multimodal transportation systems: A case study in lisbon. *European Transport Research Review*, *14*. https://doi.org/10.1186/s12544-022-00552-3

BAZEB. (n.d.). Bestuursakkoord zero emissie regionaal openbaar vervoer per bus [Accessed on 15/03/2025]. https://zeroemissiebus.nl

Behnia, F., Schuelke-Leech, B., & Mirhassani, M. (2024). Optimizing sustainable urban mobility: A comprehensive review of electric bus scheduling strategies and future directions. *Sustainable Cities and Society*, *108*, 105497. https://doi.org/10.1016/j.scs.2024.105497

Belloni, M., Tarsitano, D., & Sabbioni, E. (2025). A comprehensive analysis of energy consumption in battery-electric buses using experimental data: Impact of driver behavior, route characteristics, and environmental conditions. *Electronics*, *14*, 735. https://doi.org/10.3390/electronics14040735

Bie, Y., Ji, J., Wang, X., & Qu, X. (2021). Optimization of electric bus scheduling considering stochastic volatilities in trip travel time and energy consumption. *Computer-Aided Civil and Infrastructure Engineering*, *36*, 1530–1548. https://doi.org/10.1111/mice.12684

Cats, O. (2016). The robustness value of public transport development plans. *Journal of Transport Geography*, *51*, 236–246. https://doi.org/10.1016/j.jtrangeo.2016.01.011

Cats, O., & Jenelius, E. (2015). Planning for the unexpected: The value of reserve capacity for public transport network robustness. *Transportation Research Part a Policy and Practice*, *81*, 47–61. https://doi.org/10.1016/j.tra.2015.02.013

Cats, O., Yap, M., & Oort, N. v. (2016). Exposing the role of exposure: Public transport network risk analysis. *Transportation Research Part a Policy and Practice*, *88*, 1–14. https://doi.org/10.1016/j.tra.2016.03.015

Chau, M. L., Koutsompina, D., & Gkiotsalitis, K. (2024). The electric vehicle scheduling problem for buses in networks with multi-port charging stations. *Sustainability*, *16*, 1305. https://doi.org/10.3390/su16031305

Chen, Y., Zhang, Y., & Sun, R. (2021). Data-driven estimation of energy consumption for electric bus under real-world driving conditions. *Transportation Research Part D: Transport and Environment*, *98*, 102969. https://doi.org/10.1016/j.trd.2021.102969

Cicchini, T., Caridi, I., & Ermann, L. (2024). Robustness of the public transport network against attacks on its routes. *Chaos Solitons & Fractals*, *184*, 115019. https://doi.org/10.1016/j.chaos.2024.115019

CROW-KpVV. (2026, October). Monitor zero-emissiebussen nederland [Versie 1 oktober 2026].

CROW-KpVV & Ecorys. (2021, August). *Kostenkengetallen regionaal openbaar vervoer 2021* (Rapport) (Artikelnummer K-D124). CROW. Ede.

De Bruin, P., Van den Akker, M., Hoogeveen, H., & Van Kooten Niekerk, M. (2023). Scheduling electric buses with stochastic driving times. *23rd Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2023)*, *115*, 14:1–14:19. https://doi.org/10.4230/OASIcs.ATMOS.2023.14

Desaulniers, G., & Hickman, M. (2007). Chapter 2 public transit. *Handbooks in Operations Research and Management Science*, 69–127. https://doi.org/10.1016/s0927-0507(06)14002-5

DOVA. (2022, October). *Methode berekening LBI (Landelijke Bijdrage Index)* (Methodedocument, vastgesteld door het ROVB op 13 oktober 2022). Samenwerkingsverband van decentrale OV-autoriteiten (DOVA).

DOVA. (2025, December). *LBI 2025 definitief en LBI prognose 2026* (Rapport). Samenwerkingsverband van decentrale OV-autoriteiten (DOVA).

Edwards, A. J. (2022). Aggressive-dynamics metrics for drive-cycle characterization. *Transportation Research Interdisciplinary Perspectives*, *14*, 100592. https://doi.org/10.1016/j.trip.2022.100592

European Commission. (2024). Intercity electric buses: Current landscape and future prospects in europe [Accessed on 15/03/2025]. https://alternative-fuels-observatory.ec.europa.eu/general-information/news/intercity-electric-buses-current-landscape-and-future-prospects-europe

European Parlement. (2024). Greenhouse gas emissions by country and sector (infographic) [Accessed on 22/03/2025]. https://www.europarl.europa.eu/topics/en/article/20180301STO98928/greenhouse-gas-emissions-by-country-and-sector-infographic

Fan, D., Feng, Q., Zhang, A., Ren, Y., & Wang, Y. (2024). Optimization of scheduling and timetabling for multiple electric bus lines considering nonlinear energy consumption model. *Ieee Transactions on Intelligent Transportation Systems*, *25*, 5342–5355. https://doi.org/10.1109/tits.2023.3336326

Feng, F., Bao, S., Sayer, J. R., Flannagan, C., Manser, M., & Wunderlich, R. (2017). Can vehicle longitudinal jerk be used to identify aggressive drivers? an examination using naturalistic driving data. *Accident Analysis & Prevention*, *104*, 125–136. https://doi.org/10.1016/j.aap.2017.04.012

Gao, W., Lu, S., Liu, K., & Li, D. (2025). Integrated optimization of timetabling and vehicle scheduling for pure electric buses considering multiple depots and vehicle types. *Computers & Industrial Engineering*, *201*, 110833. https://doi.org/10.1016/j.cie.2024.110833

Ge, L., Voß, S., & Xie, L. (2022). Robustness and disturbances in public transport. *Public Transport*, *14*, 191–261. https://doi.org/10.1007/s12469-022-00301-8

Ibarra-Rojas, O. J., Delgado, F., Giesen, R., & Muñoz, J. C. (2015). Planning, operation, and control of bus transport systems: A literature review. *Transportation Research Part B: Methodological*, *77*, 38–75. https://doi.org/10.1016/j.trb.2015.03.002

IPCC. (2021). *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change* (V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou, Eds.). Cambridge University Press. https://doi.org/10.1017/9781009157896

Jahic, A., Eskander, M., Avdevičius, E., & Schulz, D. (2023). Energy consumption of battery- electric buses: Review of influential parameters and modelling approaches. *B&h Electrical Engineering*, *17*, 7–17. https://doi.org/10.2478/bhee-2023-0007

Jenelius, E., & Cats, O. (2015). The value of new public transport links for network robustness and redundancy. *Transportmetrica a Transport Science*, *11*, 819–835. https://doi.org/10.1080/23249935.2015.1087232

Ji, J., Bie, Y., Zeng, Z., & Wang, L. (2022). Trip energy consumption estimation for electric buses. *Communications in Transportation Research*, *2*, 100069. https://doi.org/10.1016/j.commtr.2022.100069

Karekla, X., Gkiotsalitis, K., & Tyler, N. (2020). The impact of a passenger-safety-driven acceleration limit on the operation of a bus service. *Accident Analysis & Prevention*, *148*, 105790. https://doi.org/10.1016/j.aap.2020.105790

Knoope, M. (2023). *Nieuwe waarderingskengetallen voor reistijd, betrouwbaarheid en comfort* (tech. rep.). Kennisinstituut voor Mobiliteitsbeleid.

Li, P., Zhang, Y., Zhang, Y., Zhang, K., & Jiang, M. (2021). The effects of dynamic traffic conditions, route characteristics and environmental conditions on trip-based electricity consumption prediction of electric bus. *Energy*, *218*, 119437. https://doi.org/10.1016/j.energy.2020.119437

Lim, L. K., Muis, Z. A., Ho, W. S., Hashim, H., & Bong, C. P. C. (2023). Review of the energy forecasting and scheduling model for electric buses. *Energy*, *263*, 125773. https://doi.org/10.1016/j.energy.2022.125773

Lusby, R. M., Larsen, J., & Bull, S. H. (2018). A survey on robustness in railway planning. *European Journal of Operational Research*, *266*, 1–15. https://doi.org/10.1016/j.ejor.2017.07.044

Manzolli, J. A., Trovão, J. P. F., & Antunes, C. H. (2025). Electric bus fleet charging management: A robust optimisation framework addressing battery ageing, time-of-use tariffs, and energy consumption uncertainty. *[journal to be confirmed]*.

Massoud, R., Bellotti, F., Berta, R., De Gloria, A., & Poslad, S. (2019). Eco-driving profiling and behavioral shifts using iot vehicular sensors combined with serious games. *2019 IEEE Conference on Games (CoG)*, 1–8. https://doi.org/10.1109/CIG.2019.8847992

Momen, S., Maknoon, Y., van Arem, B., & Azadeh, S. S. (2025). Designing a robust and cost-efficient electrified bus network with sparse energy consumption data. *Transportation Research Part C Emerging Technologies*, *171*, 105020. https://doi.org/10.1016/j.trc.2025.105020

Nan, S., Liao, F., Li, T., Chen, H., & Sun, J. (2023). Identifying the electricity-saving driving behaviors of electric bus based on trip-level electricity consumption: A machine learning approach. *Environmental Science and Pollution Research*, *30*, 82743–82759. https://doi.org/10.1007/s11356-023-28107-6

Nan, S., Tu, R., Li, T., Sun, J., & Chen, H. (2022). From driving behavior to energy consumption: A novel method to predict the energy consumption of electric bus. *Energy*, *261*, 125188. https://doi.org/10.1016/j.energy.2022.125188

Palk, M., Voß, P. S., & Jovanović, R. (2026). Robustness assessment of public transport networks in various graph representations: Systematic review, decision support, and case study. *Networks*, *87*, 343–371. https://doi.org/10.1002/net.70021

Peek, G., & Hagen, M. v. (2002). Creating synergy in and around stations: Three strategies for adding value. *Transportation Research Record: Journal of the Transportation Research Board*, *1793*, 1–6. https://doi.org/10.3141/1793-01

Perumal, S. S. G. R., Lusby, R. M., & Larsen, J. (2022). Electric bus planning & scheduling: A review of related problems and methodologies. *European Journal of Operational Research*, *301*, 395–413. https://doi.org/10.1016/j.ejor.2021.10.058

Rijkswaterstaat. (2026). Factsheet: Verbeteren ov-verbindingen [Accessed on 20/11/2026].

Szczurowski, J., Lubecki, A., Bałys, M., Brodawka, E., & Katarzyna, Z. (2022). Life cycle assessment study on the public transport bus fleet electrification in the context of sustainable urban development strategy. *Science of the Total Environment*, *824*, 153872. https://doi.org/10.1016/j.scitotenv.2022.153872

Tang, X., Lin, X., & He, F. (2019). Robust scheduling strategies of electric buses under stochastic traffic conditions. *Transportation Research Part C: Emerging Technologies*, *105*, 163–182. https://doi.org/10.1016/j.trc.2019.05.032

van Dorland, R., Beersma, J., Bessembinder, J., Bloemendaal, N., van den Brink, H., Brotons Blanes, M., Drijfhout, S., Haarsma, R., Keizer, I., Krikken, F., Le Bars, D., Lenderink, G., van Meijgaard, E., Meirink, J. F., Reerink, T., Selten, F., Severijns, C., Siegmund, P., Sterl, A., … van der Wiel, K. (2023). *Knmi national climate scenarios 2023 for the netherlands* (Scientific report No. WR-23-02). Royal Netherlands Meteorological Institute (KNMI).

van Engelen, A. F. V., & Geurts, H. A. M. (1983). *Een rekenmodel dat het verloop van de temperatuur over een etmaal berekent uit drie termijnmetingen van de temperatuur* (Historische Weerkundige Waarnemingen / KNMI-publicatie No. 165-III). Koninklijk Nederlands Meteorologisch Instituut (KNMI). De Bilt.

Van Oort, N. (2011). *Service reliability and urban public transport design* [Doctoral dissertation, TU Delft].

Wang, Y., Zhang, S., Liu, L., Gong, P., Lu, W., Wu, F., Gu, J., Li, Y., & Cao, Z. (2025). A review of battery electric public transport timetabling and scheduling: A 10 year retrospective and new developments. *Electronics*, *14*, 1694. https://doi.org/10.3390/electronics14091694

Wardman, M., Chintakayala, P. K., & de Jong, G. (2016). Values of travel time in europe: Review and meta-analysis. *Transportation Research Part a Policy and Practice*, *94*, 93–111. https://doi.org/10.1016/j.tra.2016.08.019

Wen, M., Linde, E., Røpke, S., Mirchandani, P., & Larsen, A. (2016). An adaptive large neighborhood search heuristic for the electric vehicle scheduling problem. *Computers & Operations Research*, *76*, 73–83. https://doi.org/10.1016/j.cor.2016.06.013

Winkel, K. N. d., Irmak, T., Happee, R., & Shyrokau, B. (2023). Standards for passenger comfort in automated vehicles: Acceleration and jerk. *Applied Ergonomics*, *106*, 103881. https://doi.org/10.1016/j.apergo.2022.103881

Wu, W., Lin, Y., Liu, R., & Jin, W. (2022). The multi-depot electric vehicle scheduling problem with power grid characteristics. *Transportation Research Part B Methodological*, *155*, 322–347. https://doi.org/10.1016/j.trb.2021.11.007

Yao, E., Liu, T., Lu, T., & Yang, Y. (2020). Optimization of electric vehicle scheduling with multiple vehicle types in public transport. *Sustainable Cities and Society*, *52*, 101862. https://doi.org/10.1016/j.scs.2019.101862

Yap, M. (2014). *Robust public transport from a passenger perspective* [Master's thesis, TU Delft].

Yap, M., & Van Oort, N. Driver schedule efficiency vs. public transport robustness: A framework to quantify this trade-off based on passive data. In: In *Proceedings of the conference on advanced systems in public transport and transitdata (caspt 2018)*. 2018.

Yıldırım, Ş., & Yıldız, B. (2021). Electric bus fleet composition and scheduling. *Transportation Research Part C Emerging Technologies*, *129*, 103197. https://doi.org/10.1016/j.trc.2021.103197

Zhang, A., Li, T., Zheng, Y., Li, X., Abdullah, M., & Dong, C. (2022). Mixed electric bus fleet scheduling problem with partial mixed-route and partial recharging. *International Journal of Sustainable Transportation*, *16*, 73–83. https://doi.org/10.1080/15568318.2021.1914791

Zhang, L., Wang, S., & Qu, X. (2021). Optimal electric bus fleet scheduling considering battery degradation and non-linear charging profile. *Transportation Research Part E Logistics and Transportation Review*, *154*, 102445. https://doi.org/10.1016/j.tre.2021.102445

Zhang, L., Han, Y., Peng, J., & Wang, Y. (2023). Vehicle and charging scheduling of electric bus fleets: A comprehensive review. *Journal of Intelligent and Connected Vehicles*, *6*(3), 116–124. https://doi.org/10.26599/JICV.2023.9210012

Zhang, W., Fu, B., Zhou, G., Song, Z., Xi, C., & Li, D. (2024). Measuring carbon emissions throughout the lifecycles of urban conventional buses. *Energy Science & Engineering*, *12*, 1298–1316. https://doi.org/10.1002/ese3.1668

Zhang, X., Zhang, Z., Liu, Y., Xu, Z., & Qu, X. (2024). A review of machine learning approaches for electric vehicle energy consumption modelling in urban transportation. *Renewable Energy*, *234*, 121243. https://doi.org/10.1016/j.renene.2024.121243

Zhang, Z., Ye, B., Wang, S., & Ma, Y. (2024). Analysis and estimation of energy consumption of electric buses using real-world data. *Transportation Research Part D: Transport and Environment*, *126*, 104017. https://doi.org/10.1016/j.trd.2023.104017

A

Vehicle tasks and bus changes

The validation of the EC model on vehicle task level (section 6.1.4) requires the trips of a vehicle task to be collected from the data. This is complicated by trips that are not recorded and by bus changes at the start of, or during, the day. Figures A.1 and A.2 show, for every vehicle task of the Apeldoorn and Lelystad depot and for every day of 2025, how many times the bus was changed during the task. A white cell means that the vehicle task could not be found in the data on that day, for example because the trips were not recorded. The bus changes are capped at four in the figures.

The figures show that bus changes are the rule rather than the exception for a large part of the vehicle tasks. In Apeldoorn, most vehicle tasks have one or two bus changes on many days. In Lelystad, several vehicle tasks (for example 4505 and 4507) have two or more bus changes on a large share of the days in the first half of the year, whereas other tasks are mostly driven by one or two vehicles. The pattern in Lelystad also changes over the year, with fewer changes from the summer onward. A likely reason for the changes is that buses are replaced when the SoC is depleting faster than expected, but the data does not state the reason. Because of these changes, the validation does not group the trips by planned vehicle task, but by unique vehicle and day, as explained in section 6.1.4.

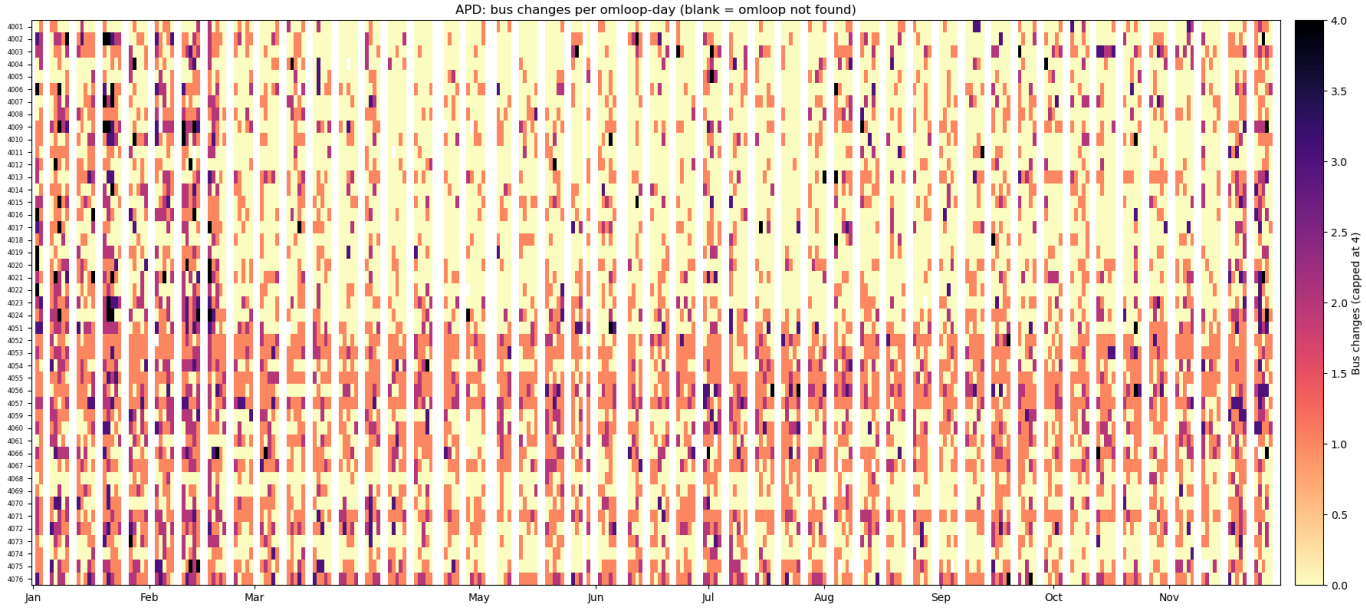


Figure A.1: Bus changes per vehicle task and day for the Apeldoorn depot (blank = vehicle task not found in the data)

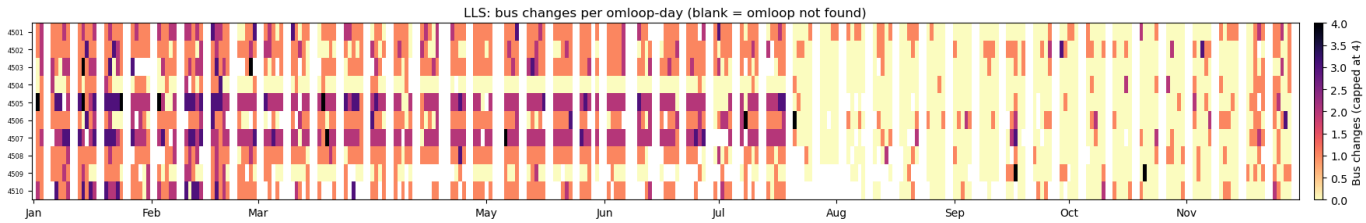


Figure A.2: Bus changes per vehicle task and day for the Lelystad depot (blank = vehicle task not found in the data)

B

Speed dynamics as additional factor

This appendix describes the exploratory analysis of an indicator for speed dynamics, as introduced in section 3.1.1. The analysis was carried out in an early stage of the research. It did not lead to an additional factor in the EC model, because the data requirements are too high for the application to a full year of trips. It is included for the sake of completeness and as a starting point for future research.

B.1. Speed dynamics indicators

Previous studies have explored the impact of driving behaviour on energy consumption by using different indicators based on speed, acceleration and/or the position of the gas and brake pedal to describe driving aggressiveness (see for example Abdelaty et al. (2021), Nan et al. (2022, 2023), and Z. Zhang et al. (2024)). Because speed profiles can be influenced by more than just driving behaviour (e.g. traffic conditions or the number of stops for boarding), a broader definition is used in the form of speed dynamics. The CANbus system of the bus collects speed records up to multiple times per second. These registrations give a speed profile for the trip. From the speed profile five indicators are derived that together give a relatively complete picture of speed dynamics (Edwards, 2022; Feng et al., 2017; Massoud et al., 2019). All five indicators are normalised and combined to form a single continuous variable. The normalisation is done across all trips for a global score. In addition, a line-specific score is normalised on all the trips of a single line.

Positive kinetic energy (PKE) explains the total positive kinetic energy provided by the bus per unit of distance.

$$\text{PKE} = \frac{\sum_{i=1}^{n} \max(\Delta v_i, 0)^2}{d} \tag{B.1}$$

where $\Delta v_i = v_i - v_{i-1}$ is the change in speed (m/s) and d the total distance (km).

Relative positive acceleration (RPA) is the amount of acceleration weighted by the speed, summed over the total trip distance. It tells something about the average magnitude of acceleration.

$$\text{RPA} = \frac{\sum_{i=1}^{n} a_i^+ \cdot v_i \cdot \Delta t_i}{\sum_{i=1}^{n} v_i \cdot \Delta t_i} \tag{B.2}$$

where a_i^+ is the positive acceleration at time step i (m/s^2), v_i the speed (m/s), and Δt_i the time interval (s).

Hard acceleration/braking events per km (HAE/HBE): uncomfortable acceleration and braking rates can tell something about the anticipation of traffic events. The threshold values are based on Winkel et al. (2023) and Karekla et al. (2020).

$$\text{HAE} = \frac{\#\{i : a_i > a^+_{\text{threshold}}\}}{d}, \qquad \text{HBE} = \frac{\#\{i : a_i < a^-_{\text{threshold}}\}}{d} \tag{B.3}$$

where $a^{+}_{\text{threshold}} = 1.0$ m/s^2, $a^{-}_{\text{threshold}} = -1.0$ m/s^2 and d the total distance (km).

RMS jerk describes the smoothness of the acceleration. It illustrates the rate at which the acceleration of the bus changes during a unit of time. The indicator becomes higher when accelerating is abruptly followed by braking.

$$\text{RMS}_{\text{jerk}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} j_i^2} \tag{B.4}$$

where $j_i = \dfrac{a_i - a_{i-1}}{\Delta t_i}$ is the jerk at time step i (m/s^3).

B.2. Data preparation

To generate data that can quantify speed dynamics, speed profiles are used. A speed profile is a graph that contains the vehicle's speed on the Y-axis and time on the X-axis. The graph can be generated by having a large number of speed measurements at preferably equal intervals. EBS has this data available on trip level. The time intervals between measurements range from multiple data points per second to intervals of more than 10 seconds. With multiple data points per second, gathering data from a day of trips results in large files. Therefore, the data set that is used to test the influence of driving style is capped at 833 trips, which already contain over 2,800,000 speed points. Not all trips were valid, however, and roughly 30% of the speed measurements were duplicates. After filtering trips on a minimum duration of 10 minutes, an average speed between 10 and 60 km/h and no time intervals larger than 20 seconds, 601 trips remained.

All time intervals were then set to 1 Hz. Gaps were filled with linear interpolation, and when multiple speeds were recorded in the same second, the mean value was taken. Acceleration rates higher than 3.5 m/s^2 or lower than -5.5 m/s^2 were smoothened, as they reflect unrealistic vehicle behaviour. Speed profiles with long idle times at the start or the end of a trip had that part trimmed off. The speed profiles with the highest and lowest global and line-specific scores are presented in figure B.1. The scores are normalised between all available records, so the global and line-specific score are relative. The results of each filter and the distribution of the scores are presented in figure B.2.

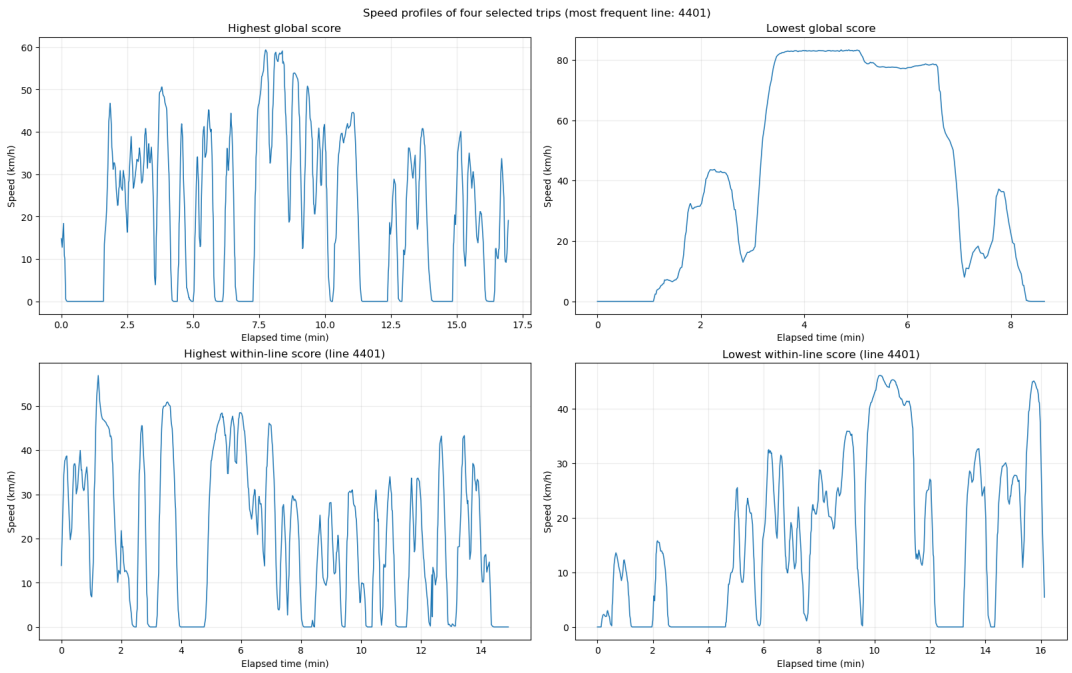


Figure B.1: Speed profiles of trips with the most extreme global and line-specific scores

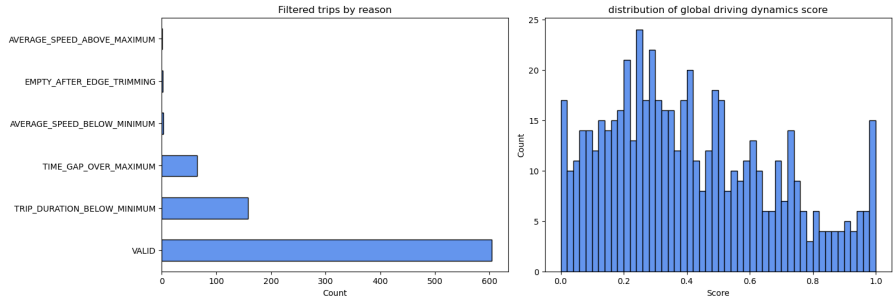


Figure B.2: Filter results and distribution of the global score for the speed dynamics indicator

B.3. Effect on the EC estimation

To implement the speed dynamics scores in the regression model, one additional filter had to be applied to align with the base model: a minimum distance of 20 km. This decreased the number of trips to 125, a fraction of the data set that was used to train the base model. The smaller sample size may lead to over-fitting, certainly when applying a non-linear regression model. Having 7 factors means that there are 21 features in the model, compared to 125 measurements. Because of the small sample size, only the global score is included, as it should be able to capture the difference between all trips. The line-specific score would need more trips to gain a reliable metric for all lines. To prevent over-fitting, ridge regression is used instead of ordinary least squares, as it penalises features that contribute little to the model. Furthermore, a linear model is used. While the most influential factor, temperature, was shown to have a non-linear relationship, the fact that the trips are from a single day limits the range in which the values occur, and thus the consequences of assuming a linear relationship.

The scatter plots in figure B.3 show the estimation results for the base model without the speed dynamics score and for the model with the score. Both models have a low predictive value, which can be explained by the small sample size and the already low variance in EC rates within the data set. With leave-one-out cross-validation, the R^2 increases from 0.21 to 0.31 when the score is added, so the addition of speed dynamics does improve the model's ability to estimate EC. Figure B.4 shows that the global score is the most important feature when it is included. Because a linear model is used here, the highest-ranking features also translate into the most important factors.

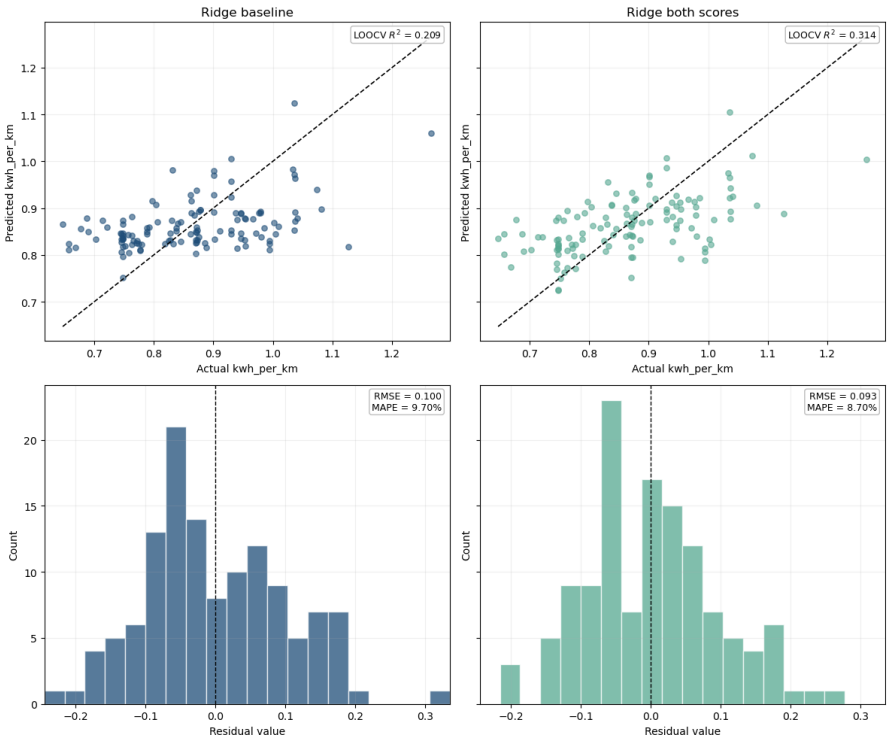


Figure B.3: Comparison between the ridge regression base model without the speed dynamics score (left), and the model with the score (right)

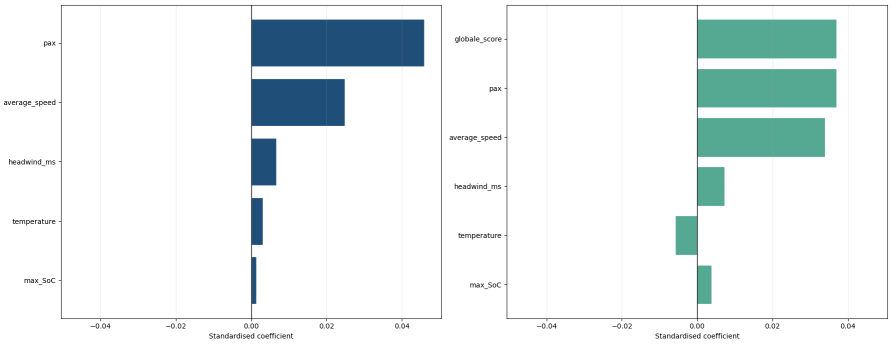


Figure B.4: Importance of the features in the ridge regression model with the speed dynamics score

B.4. Conclusion

The global speed dynamics score is worth exploring further. However, the speed data of only a single day can already be larger than the trip data of a whole year (500 MB or more), which prevents obtaining a reliable model with a sufficient number of trips. The score is therefore not included in the remainder of this research. As the speed dynamics are not used in the simulation, the outcomes of this research are not affected by this choice. Obtaining the speed data for a larger number of trips, or at a lower resolution, would be a logical step for future research.

C

Results without correction of the benchmark

In the sensitivity analysis (section 6.4.1) the planned schedule cost of the alternative schedules is compared with a reference, the average of the -10 and $+10$ schedules, instead of with the planned schedule cost of the benchmark schedule. This was done because the benchmark schedule is less efficient than the $+10$ schedule, which does not fit the trend of the other schedules, and because comparing with the benchmark would give too positive outcomes for more energy redundancy. For transparency, table C.1 shows the planned schedule cost and the overall net change in cost when the planned schedule cost of the benchmark itself is used. The cancellations and the passenger delay are the same as in the main analysis. All values are percentages of the annual operator cost of the benchmark schedule without driver cost, as in table 6.12 and table 6.16.

The comparison shows that the correction mainly affects the schedules with more energy redundancy. Without the correction, the planned cost of the $+10$ and $+20$ schedules in Lelystad is -0.66% and $+0.32\%$ relative to the benchmark, instead of $+0.33\%$ and $+1.31\%$ relative to the reference, so the overall savings are larger (-3.10% and -4.50% instead of -2.11% and -3.51%). With driver cost the effect is larger, as the benchmark schedule has many more idle and dead-heading hours than the $+10$ schedule. The $+10$ schedule would then lead to a saving of 7.8% in Lelystad and 3.4% in Apeldoorn. These numbers include the difference in efficiency between the benchmark and the alternative schedules that is not caused by the EC value, and should therefore not be used to judge the effect of the energy redundancy.

Table C.1: Planned schedule cost and overall net change in cost when the planned schedule cost of the benchmark is used as reference (without correction), as a percentage of the annual operator cost of the benchmark without driver cost

	Without driver cost		With driver cost	
Schedule	**Planned**	**Overall**	**Planned**	**Overall**
Apeldoorn				
-20	$+0.00$	$+7.28$	-3.24	$+4.04$
-10	-0.00	$+3.15$	-3.13	$+0.01$
$+10$	-0.02	-0.55	-2.83	-3.36
$+20$	$+1.22$	$+0.12$	$+0.21$	-0.89
Lelystad				
-20	-1.60	$+18.00$	-15.71	$+3.89$
-10	-1.32	$+10.74$	-13.45	-1.38
$+10$	-0.66	-3.10	-5.37	-7.82
$+20$	$+0.32$	-4.50	$+2.63$	-2.19

D

AI statement

AI, and more specifically LLMs are used for four purposes:

1. Coding for data analysis. As LLMs have the ability to write code with explanatory comments, a lot of programming work can be outsourced. The programme knows which data analysis techniques to apply and helps with implementing ideas or solving errors.
2. Providing LaTeX code for formulas and tables. A lot of administrative work is saved by asking AI to generate code based on code output, screenshots or Excel snippets.
3. Discussing ideas. Although AI is generally bad at providing output with scientific references, it still manages to prompt useful ideas for problems encountered during the thesis. It provides a topic or starting point to search for scientific literature.
4. Checks on text structure and flow. With the ideas typed out, an agent was asked to point out inconsistencies between sections. It also suggested improvements to better link sections and create a functional story between the chapters.

For this study, GitHub CoPilot Pro - provided by the TU Delft through an educational licence - is mainly used. Next to public data, such as the weather data from the KNMI, only fictional data was shared to prevent leaks. For occasional questions that did not involve sensitive information (e.g. providing LaTeX code), Google's Gemini and Antrophic's Claude were consulted. Output from the models was always verified.