Research Article

Transportation Research Record
2023, Vol. 2677(3) 389–401

DOI: 10.1177/03611981221115427
journals.sagepub.com/home/trr

Big Data and Discrete Optimization for Electric Urban Bus Operations

Samuel Würtz[1], Klaus Bogenberger[2], and Ulrich Göhner[3]

Abstract

The electrification of urban bus fleets is a challenging task, especially for smaller public transport operators. The main challenge lies in the uncertainty about many technical aspects, like range of vehicles under different circumstances or charging times, that are new for the operators. The purpose of this research is to introduce an approach to solve this problem by incorporating all available data from an existing bus fleet and finding an optimal solution with discrete mathematical optimization. Extensive data logging in the project enabled us to leverage tracking data from the whole bus network including trajectories, powertrain data, and operational data. This enabled us to validate assumptions about the energy demand, waiting times, and different traffic situations during the day. To get better insights into the requirements of an urban bus fleet, we simulated the potential electric buses in detail and extracted other necessary data like actual dwell times. Based on the simulation results and processed data, we implemented a linear programming model to search for a cost-optimal configuration of vehicles and charging infrastructure. We tested the framework with a scenario in which we analyzed the solutions with different numbers of diesel buses in the fleet. The application of our algorithm shows that it can produce optimal results in a short amount of time, for a medium-sized city in Germany. We also demonstrate that the flexible and constraint-based formulation of this approach allows it to be incorporated in the planning process of most public transport operators.

Keywords

big data analytics, public transportation optimization, public transportation, transportation and sustainability, electric and hybrid-electric vehicles

Political regulations have led to rising pressure on public transport providers to reduce local emissions by transitioning their whole system from diesel-fueled buses to emission-free vehicles. Clean vehicles in that sense have complete, or at least partial, emission-free driving (hydrogen, battery-electric or hybrid). This is formulated by the European Commission (EC) as the Green Vehicle Directive, which imposes stepwise higher percentage rates of emission-free engines. From 2026, 65% of newly purchased heavy-duty vehicles must be operated emission free (*1*). This is a big challenge for many public transport operators (PTOs), especially smaller ones, since electric buses have totally different technical requirements, and their operation is limited by additional constraints. But this also presents a great opportunity. Electric buses could pioneer a new age of clean and efficient urban transport and put cities on track toward sustainability (*2*). Some big cities like Shenzhen, China, have adapted a large proportion of their bus fleet to electric drives, but other cities have been more hesitant. Better incentives and higher planning reliability could help the adoption of electric buses.

The amount of data collected about travel behavior of people, from vehicles, and about the weather is growing vastly and this trend is expected to accelerate even more (*3*). Considering this, it seems natural to leverage this data for new applications. We propose a framework for accurately evaluating and optimizing the process of electrifying a bus fleet with battery-electric vehicles. We

[1]Allgäu Research Centre, University of Applied Sciences Kempten, Kempten, Germany
[2]TUM Chair of Traffic Engineering and Control, Technical University of Munich, Germany
[3]Department of Computer Science, University of Applied Sciences Kempten, Kempten, Germany

Corresponding Author:
Samuel Würtz, samuel.wuertz@hs-kempten.de

collected and processed the required data from an urban bus fleet and formulated a mathematical model to optimize the choice of vehicles, and number and locations of charging stations for the bus network. One main goal was to make the model adjustable to custom constraints to be relevant in real-world scenarios, easy to integrate into the regular planning process undertaken by any public transport operator (PTO). During the planning phase the PTOs have various tasks in mind, for example trying to maximize vehicle utilization and minimize labor cost. They also need to know suitable spots for lunch breaks and parking lots for the buses to create feasible bus routes. Since it will be very challenging to create a system that covers all possibilities, we decided to leave the plan data (in the following referred to as rotation plans) untouched and focus on finding optimal solutions for the given bus routes, thus achieving higher applicability. The proposed framework includes evaluating the potential of electrification with different vehicles and charging infrastructure as well as analyzing a stepwise adoption of electric buses, allowing for some diesel buses in the starting phase.

The rest of the paper will be structured as follows: literature review, methodology, case study, results, and conclusion. The literature review serves as a quick glance at past work in the field of electrifying urban buses and formulates the strengths of our approach. In the methodology part, we explain the general framework consisting of the data collection, preprocessing, and the mathematical model itself. Then, a section describes the use case area where the project is conducted. Results are shown from using the framework on our use case. In the conclusion, we summarize the findings and give an outlook on future work.

Literature Review

The research on optimizing urban public bus transport dates back to the early 1980s. It was first described as the Urban Transit Route Network Design Problem (UTRNDP) in Ceder and Wilson (*4*), which divided the problem into five stages: route definition, frequency setting, timetable creation, vehicle scheduling, and driver scheduling. Since then, many approaches to solving those problems have been published. We do not want to include a comprehensive list, but some notable results were made by Mandl (*5*) as early as 1980. He created data sets which resemble cities in Switzerland, developed heuristic approaches on them, and published the data sets for comparison of future approaches. The more recent work of Iliopoulou et al. (*6*) and Ahmed et al. (*7*), who leveraged genetic algorithms to solve these problems, were among the first approaches that included electric buses. All those approaches included solving a variation of the vehicle routing problem to come up with optimal routes for the buses, which is an NP-hard problem.

Many of the approaches cannot include the variety of constraints which come up in real-world scenarios and, since we assume that the PTOs have reasonably good routes, we decided to focus on generating optimal solutions for charging locations and battery sizes for a fixed set of planned bus routes. Since we do not desire to change the current state of the bus network that we would like to electrify, we took another approach to the problem, as we will describe later. We oriented our goals more toward positioning of charging infrastructure and sizing of the battery of buses. Uslu and Kaya (*8*) proposed an approach for optimizing charging infrastructure for intercity electric bus networks. They setup a solution for a charging infrastructure for any vehicle fleet to use. Their optimization model was based on average energy demand and the distances between cities. Kameda and Mukai (*9*) and Sadeghi-Barzani et al. (*10*) used similar approaches to setup a cost-effective charging infrastructure based on mobility data, with the goal of a cost-efficient charging infrastructure within a city. Kunith et al. (*11*) went a step further and generated a cost-effective infrastructure setup for the specific bus network, assuming a fixed battery size. The exclusion of the battery size from the optimization might prevent more cost-effective solutions, based on smaller batteries or single vehicles with higher battery capacity to reduce the number of charging stations. There are a few approaches that combine the sizing of battery with the setup of charging points. In Gao et al. (*12*) a solution for optimizing both infrastructure and battery size, by considering fast charging and battery swapping based on standardized bus cycles, was proposed.

De Filippo et al. (*13*) developed an agent-based simulation, which considered energy consumption and charging times to evaluate feasibility of electric operations of a bus network. This is an interesting approach but does not help directly in generating optimal solutions for vehicle and charger choice. For a cost-efficient implementation of an electric bus fleet, it is important to consider the battery capacity and energy demand of the vehicles as well as the sizing and location of charging stations. Jefferies and Göhlich (*14*), Kunith et al. (*18*), and Berthold (*15*) developed different approaches based on mixed-integer linear programming (MILP) models. Berthold (*15*) had some tracking data for single bus lines available and used that to incorporate some real-world data, whereas Kunith et al. (*18*) and Jefferies and Göhlich (*14*) had to take some simplifying assumptions to estimate the energy demand. For a more in-depth look into the developments in that field, we recommend the work of Jefferies and Göhlich (*14*), which gives a very

Table 1. Comparison of related works

Research work	Charger location	Battery size	Simplified energy demand	Drive cycle energy demand	Topology	Passenger load	Temperature	Traffic	Route optimization	Data input type
Uslu and Kaya (*8*)	x	na	x	na	na	na	na	na	na	plan
Kameda and Mukai (9)	x	na	na	na	na	na	na	na	na	fcd
Sadeghi-Barzani et al. (*10*)	x	na	na	na	na	na	na	na	na	infra
Kunith et al. (*11*)	x	na	na	na	na	na	na	na	na	plan
Gao et al. (*12*)	x	x	na	x	na	na	na	na	na	cycle
De Filippo et al. (*13*)	na	na	x	na	na	na	na	na	na	plan
Jefferies and Göhlich (*14*)	x	x	na	na	x	na	x	na	x	plan
Berthold (*15*)	x	x	na	na	na	na	x	x	na	tracking
Kunith (*16*)	x	x	na	x	na	na	x	na	x	plan
Rogge (*17*)	x	x	x	na	x	na	x	na	x	plan
This paper	x	x	na	na	x	x	x	x	na	tracking

Note: We distinguish between five types of data input: plan (bus schedules), fcd (car/taxi movements in the study area), infra (information about power supply infrastructure), cycle (standard driving cycles in the area), tracking (detailed tracking data for all or some of the buses in the study area); na = not applicable.

good overview. The challenge is approached from many different angles and with different scopes in mind. For the reliable planning of electric vehicles, more data of external factors is required. The topology and weather conditions are the most obvious ones, but also the state of the traffic flow affects the range of battery-electric vehicles, as is shown for example in the research of Kessler and Bogenberger (*19*), and Morlock et al. (*20*).

We developed our own mathematical model for jointly optimizing sizing and location of charging infrastructure as well as the battery size of the vehicles. The main goal was to design the model in a way that incorporated the vast amount of real-world tracking data into the system and make it flexible enough to apply it to any existing bus network. Such a system reduces the uncertainties concerning the real-world application of the algorithm. It also reduces the introduction barrier for most PTOs since they could use their accustomed tools for planning bus operations and checking for cost and feasibility of the planned approach. Table 1 compares related research, distinguishing between five types of data input.

Methodology

This section will give a detailed overview of the different stages of our proposed framework. As shown in Figure 1, the process consists of various distinct steps: (1) collection of required data, (2) evaluating and preprocessing the data for the optimization, and (3) definition of the task as a mathematical optimization model which will produce the needed result set. For a better understanding of the interconnections between the different stages, Table 2 showcases the required input and processing steps for the optimization model. The first column describes the step within the framework. The second column indicates the required input from previous stages, and the last column describes the output of the respective stage.

Data Sources

The input data consists of five categories: topological data, tracking data, the street network, timetable data, and the cost and technical data about the buses and infrastructure. We obtained the topological data from the NASA Earth Observation data set (*21*) via a Python interface (https://pypi.org/project/elevation/). This gives very accurate values with a resolution of 30 m × 30 m. For the street network, we used OSM maps (*22*), which are fairly accurate for our studied area, especially for the main roads used by urban buses.

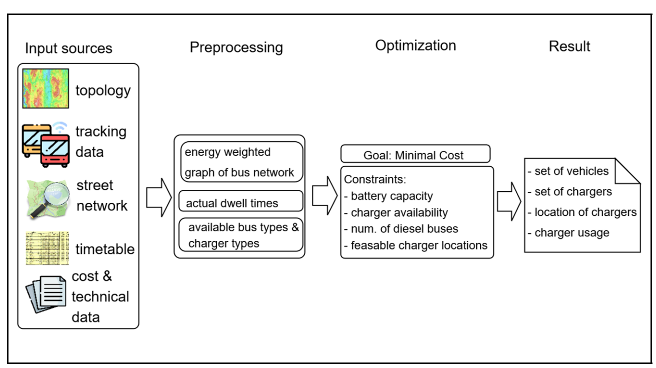

Figure 1. System overview. Workflow from input, preprocessing, optimization to the obtained results.

Table 2. Information flow through the different stages of the optimization framework

Stage	Input	Result
1. Input data		
1.1 FMS	na	1.1.1 Vehicle weight
		1.1.2 Open/close door status
		1.1.3 Ambient temperature
1.2 GPS		1.2.1 Trajectories
1.3 Technical data	na	1.3.1 Engine power, efficiency, ...
		1.3.2 Vehicle cost
		1.3.3 Battery sizes
		1.3.4 Charger powers
		1.3.5 Charger costs
1.4 Operational data	na	1.4.1 Timetables
1.5 Map data		1.5.1 Street network
		1.5.2 Bus stop locations
		1.5.3 Topology
2. Preprocessing		
2.1 Enrichment	1.1.1, 1.1.2, 1.2.1, 1.4.1, 1.5.1, 1.5.2	2.1.1 Split trajectories at bus stops
		2.1.2 Enrich trajectories with temperature and weight
		2.1.3 Calculate 10th percentile of delay and dwell times
2.2 Energy model	1.3.1, 1.3.3, 1.3.4, 1.5.3, 2.1.1, 2.1.2	2.2.1 Energy demand for every subsection
3. Optimization	1.3.2, 1.3.5, 1.4.1, 2.2.1	3.1 Type of vehicles for every circulation
		3.2 Locations of chargers
		3.3 Charger type for every location
		3.4 Usage of every charger

Note: FMS = Fleet Managment System; GPS = global positioning system; na = not applicable.
The types of input data have completely different characteristics: the FMS data is sampled with up to 100 Hz, which produces vast amounts of data; the GPS values are sampled with 1 Hz and need to get matched to the FMS data by time; the rest of the input data (technical/operational data and map data) need to be obtained once for every project.

To receive the tracking data, we equipped 50 of the 95 buses with telematic devices and were thus able to collect all the required data. These data loggers are Linux-based devices with our own logging software that transfers the data in real-time to our server infrastructure. Since the amount of data that needed to be obtained and interpreted was very large, we decided to set up a scalable Apache Kafka Cluster (*23*), which receives the data from the different buses and interprets the raw byte stream into human-readable values before storing them into a database. For efficient querying of the data, we chose a PostgreSQL Timescale database (*24*), which is optimized for large amounts of time series data. Since some powertrain specific data is sampled with 100 Hz, we must process around 60 to 100 GB of data, which results in more than 200 million data-points every day. To record all this data, we used the FMS Fleet Managment System Standard Interface (https://www.fms-standard.com/Bus/index.htm) interface of the buses, which is a standardized protocol defined in HDEI-BCEI-Task-Force (*25*) for accessing different values from the vehicle. This enables us to use the same logging software on all the different vehicles from the fleet. Therefore, we have the values of the door-open/closed status as well as ambient temperature and vehicle weight all the time for every vehicle and can incorporate it into the energy calculation. To make sense of the recorded global positioning system (GPS) tracks, we needed the plan data and to assign each track to one of the bus routes. To incorporate the plan data, we implemented the standardized interface from the VDV (Verband Deutscher Verkehrsunternehmen), which allows easy integration of the data from any bus operator that uses this interface (*26*).

For the energetic simulation of the buses and charging processes, we required technical data about the buses. Some of the more general vehicle data was freely available online, whereas other values were only directly obtainable through the manufacturer. Since our optimization goal was cost efficiency, we also needed the costs of potential vehicles and chargers. For our purposes, approximate values were enough, and we obtained them from our project partner (Göttinger Verkehrsbetriebe), the PTO for Göttingen. Because we wanted to optimize with many possible vehicles, we had to estimate costs for different battery sizes and charging powers. Those costs only served as an example and the results in this study should be taken as a proof of concept. For detailed results, one would need to investigate the costs more specifically and provide them for the optimization.

Data Processing

The purpose of the preprocessing step is to make sense of the vast amount of data and prepare it for the integer linear programming (ILP) model. Therefore, we implemented a data pipeline that generates the necessary input for the optimization algorithm. This generated input data set consists of the route lists with the actual waiting times, the energy demand for every subsection of the bus network, a set of available bus types, and a set of available charger types. The route lists were generated from the plan data in combination with the tracking data. The plan data was used to obtain a list of consecutive bus stops. One list of bus stops represents a rotation plan that one bus must take for an entire day. Those lists were also used to assign the recorded GPS tracks to a rotation plan, by a custom-made algorithm. The first step is to get the data for one bus for the whole day from the database. To assign this track to one of the tracks from a rotation plan we used the mentioned list of bus stops with the schedule time and a time window of 5 min too early until 25 min delay. If the bus visits all the bus stops from the rotation plan within the given time windows, we assign the GPS track to the rotation plan. This would be an ideal way to identify the rotation plans in the GPS data, but in reality the vehicles do not stick exactly to the plan all the time. Sometimes they skip some of the bus stops, which mostly is because of temporary detours to avoid road construction sites, accidents, or other road obstructions.

From those identified GPS tracks, we extracted the actual waiting times at each bus stop. To obtain reliable results, we used the proximity to the bus stop in combination with the speed and door status. Figure 2 depicts the decision algorithm for how the dwell times get extracted. The main decision criteria are speed and proximity to the bus stops. The values show the speed in m/s, the color shows the proximity to the bus stops, the first half of the plot shows the distance to the starting bus stop, and the second half of the plot the distance to the destination bus stop. When the bus is not moving within 50 m of the corresponding bus stops, we count this as dwell times. The cutoff time for the dwell time is marked with a vertical line in Figure 2. To validate our results, we compared this with the opening and closing of the doors. The information was obtained directly from the vehicle via the FMS-Bus, which is a standard protocol defined in HDEI-BCEI-Task-Force (*25*) for heavy-duty vehicles that provides several signals and sensory data from the vehicle. This gives reliable results for the waiting times. Because of varying traffic and in general the transit service variability, the dwell times depend heavily on the delay of the buses. To account for that, we extracted the 10^{th} percentile to get the dwell times the buses have in more than 90% of the time. This ensures that the drivers should have enough time to charge for the required duration.

The lists of bus stops were then enriched by the calculated waiting times. The energy demand is also a very important aspect of the system. Since many other approaches simply use average values for the energy demand, they need to set large safety margins to ensure that the bus reaches its destination. At Kempten University of Applied Sciences, we developed a yet unpublished energy model based on Matlab and Simulink (*27*) which can simulate an electric bus based on the data we collected from existing diesel buses. The energy demand is then computed based on the trajectory data, power demand of auxiliaries, which mainly consisted of the HVAC (heating/ventilation/air-conditioning), and the total weight of the vehicle. The extensive data collection enabled us to understand the operational conditions very well and thus to calculate accurately the energy demand to be expected. To be able to cover extreme cases as well, we came up with a matrix of temperatures and additional weight. Table 3 shows the estimated relative impact on the overall energy demand. Values for the temperature and required adjustments to the temperature inside the bus, for the passengers to feel comfortable, are taken from VDV (*28*). For every entry of this matrix, we calculated the hypothetical energy demand under strained conditions such as extreme weather conditions or high travel demand, which consequently allows validation of the robustness of the solutions.

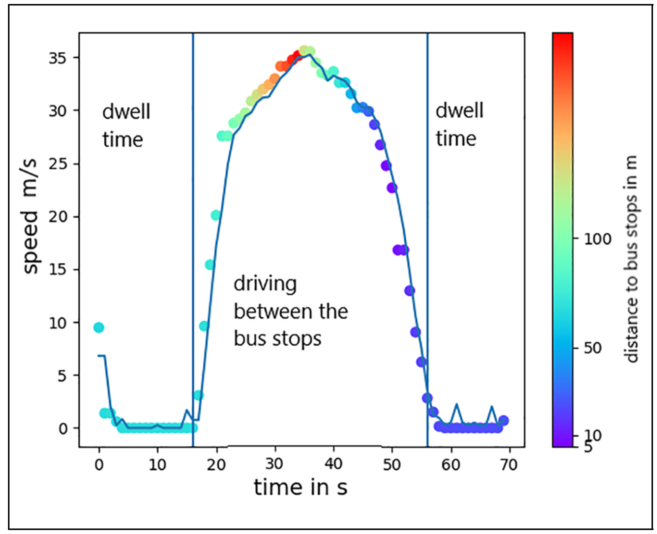

Figure 2. Example for the detection of an actual dwell time at a specific bus stop. The horizontal lines are the cutting points where we define the bus to be leaving, respectively arriving at the bus stop.

The last parts of the input data set are the available bus types and charger types. These are difficult to generate automatically, since the technical details and costs are changing at a fast pace. Therefore, these details should be investigated for every project individually. We decided to consider only the capital expenditures, as the operational

Table 3. Matrix for relative additional energy demand depending on HVAC and passenger numbers

Temperature	Passenger load		
	Empty	Average	Full
−20°C	+ + + +	+ + + + +	+ + + + + +
−10°C	+ + +	+ + + +	+ + + + +
0°C	+ +	+ + +	+ + + +
20°C	o	+	+ +
35°C	+ + + +	+ + + + +	+ + + + + +

Note: HVAC = heating/ventilation/air-conditioning.

Table 4. Input variables

Variable		Description
s	$\in \mathbb{N}$	Number of bus stop
c	$\in \mathbb{N}$	Number of charger types
r	$\in \mathbb{N}$	Number of routes
b	$\in \mathbb{N}$	Number of bus types
t	$\in \mathbb{N}$	Number of time steps
bs	$\in \mathbb{N}$	Number of all bus stops
bs_i	$i \in [r]$	Bus stops on route i
b_i	$i \in [b]$	Concrete bus from bus types
b_{b-1}	$\in \mathbb{N}$	Bus type of the diesel bus
bd	$\in \mathbb{N}$	Number of diesel buses
$D = (d_{i,j})$	$\in \mathbb{R}^{s \times s}$	Distances between bus stops
$CC = (cc_i)$	$\in \mathbb{R}^{s \times s}$	Costs of charger types
$CB = (cb_i)$	$\in \mathbb{R}^{s \times s}$	Costs of bus types
$RB = (rb_i)$	$\in \mathbb{R}^{s \times s}$	Range of bus types
$CP = (cp_i)$	$\in \mathbb{R}^{s \times s}$	Power of charger types

Table 5. Binary optimization variables

Variable		Description
$A = (\alpha_{ij})$	$\in 0, 1^{s \times c}$	1 if charger type j is used on bus stop i, 0 otherwise
$B = (\beta_{ij})$	$\in 0, 1^{b \times r}$	1 if bus type i is used on route j, 0 otherwise
$C = (\gamma_{ijkl})$	$\in 0, 1^{r \times s \times c \times t}$	1 if on route i, bus stop j, charger type k is used at timestep l, 0 otherwise

expenditures should roughly remain the same because we did not change the timetables or number of vehicles; drivers were therefore unaffected.

Optimization

The optimization problem is formulated as an ILP model. This means, we formulated an optimization goal, in our case minimal cost for buses and chargers. Additionally, we need constraints to ensure that the choice of vehicles and chargers results in a feasible solution for the PTO, in which all the bus routes can operate for one day. The ILP model was implemented in Python (*29*) using Google OR Tools (*30*). For fast and efficient solving we chose the SCIP solver which, according to benchmarks on their website (*31*), is one of the fastest non-commercial solvers for MILPas well as mixed-integer nonlinear programming (MINLP) (*32*). The SCIP solver leverages a branch-and-bound algorithm which divides the linear program (LP) in sub-problems, which are easier to solve. It builds up a tree of sub-optimization problems by limiting the domain of the different variables. If one branch has definitively worse results than the found solutions, it can be discarded and therefore speed up the solution process without missing the optimal solution. Input variables are described in Table 4 and binary optimization variables are described in Table 5.

The ILP Model is modeled to cover the entire bus network for one day of the week, therefore assuming that the buses' depot can be used for recharging overnight. The amount of energy required to charge all the buses by the next morning can also be taken from the solution set. From our investigations and meetings with the PTO, we know that the buses' batteries are only used in a state of charge (SOC) ranging between 20% and 80%. According to different studies and advice from the manufacturers in their manuals, this increases the battery life (*33–35*). A useful side effect from this is that the charging behavior in this range can be assumed to be linear (*36*), which reduces the complexity of our model. We assumed in our model that all buses use overhead charging, and this is the case for our project. Jefferies and Göhlich (*14*) also evaluated charging technologies and concluded that plugin chargers will not be feasible in practice, since it requires a lot of manual work by the driver (*14*). However, we will also investigate the possibilities of inductive charging in the future. For the time required to couple/decouple the charger, we assumed a value of 15 s taken from Jefferies and Göhlich (*14*).

We assumed a total time of 30 s at a charging stop to couple and decouple the buses from the chargers and to account for minor unforeseen events, like people or cars blocking the charging station temporarily. Based on this, we divided the problem into 1-min slices. This means that dwell times can only be considered for a charging event if it is longer than 1 min. The trade-off here is granularity versus computational complexity, since the number of time steps in the simulation greatly affects the number of variables in the model. In Equations 1–8, the ILP Model is formulated, starting with the optimization goal and then integrating all the constraints.

$$min \sum_{i=1}^{bs} \sum_{j=1}^{c} (\alpha_{ij} \times CC_j) + \sum_{i=1}^{b} \sum_{j=1}^{r} (\beta_{ij} \times CB_i) \tag{1}$$

subject to:

$$\sum_{j=1}^{bs_m} (d_{j,j-1}) - \sum_{k=1}^{c} \sum_{o=1}^{t} (\gamma_{m,i,k,o} \times CP_k) \leqslant \sum_{l=1}^{b} (\beta_{lm} \times RB_l), \quad \forall m \in \{0, \ldots, r\} \tag{2}$$

$$\sum_{j=1}^{i} (d_{j-1,j}) - \sum_{k=1}^{c} \sum_{o=1}^{t} (\gamma_{m,i,k,o} \times RC_k) \geqslant 0, \quad \forall m \in \{0, \ldots, r\}, i \in bs_m \tag{3}$$

$$\sum_{k=1}^{c} \sum_{m=1}^{r} (\gamma_{m,i,k,o}) \leqslant \sum_{k=1}^{c} (\alpha_{i,k}), \quad \forall m \in \{0, \ldots, s\}, o \in \{0, \ldots, t\} \tag{4}$$

$$\sum_{i=1}^{r} (\beta_{bi}) = 1 \tag{5}$$

optional constraints:

$$\sum_{i=1}^{r} (\beta_{b_{b-1},i}) \leqslant bd \tag{6}$$

$$\alpha_{x,y} = 1 \tag{7}$$

$$\alpha_{x,y} = 0 \tag{8}$$

Unless specified otherwise, every bus stop is a potential charging location, and every route could be served by any of the available bus types. This forms the basis for the definition of the optimization goal in Equation 1. α_{ij} denotes if a charger of type j is needed at bus stop i; then all those required chargers are multiplied by their respective cost and added up for the total charger cost. The second half of the optimization goal calculates the total cost for the buses by multiplying the decision variable β_{ij}, which holds the chosen bus type for a given route, with the cost of the bus type. To ensure that all buses can operate their routes without exceeding the limit of the battery, we introduced the route constraint in Equation 2. This sums up all the energy demands of subsections between two bus stops $d_{j-1,j}$ and compares this in the simplest case with the battery capacity of the chosen bus for that route. If a charger was necessary to manage the whole route, the charged energy is subtracted from the required energy. This is computed at any bus stop that has a charger available by multiplying the charging power per minute with the number of minutes the bus stays at the location (minus 1 minute to compensate for the coupling and decoupling of the charger). This may still lead to undesired behavior, if a bus has a long waiting time at the end of the trip but cannot reach that charger. Therefore, we had to include this constraint for every sub-tour. This means that for every route, we get exactly the same number of route constraints as the route has bus stops. This also ensures that we never go below the defined lower bound of 20% SOC. In analogy to this, Equation 3 ensures that the bus cannot charge more than the 80% SOC, by ensuring that the energy demand can never go below 0 for any given route. The charger capacity constraint formulated in Equation 4, ensures that there is never more than one bus charging at a charger at any given time. This means we need to create a constraint for every bus stop and every time step in the simulation time to ensure this. Furthermore, it is important to restrict the model, to ensure that the number of buses per route is exactly one, as defined in Equation 5. All those constraints were generated by our program based on the provided input data, which enables us to apply the optimization algorithm to new bus networks with minimal manual work, since it would be almost infeasible to formulate all the constraints manually for all the bus routes and sub-routes to the different bus stops. From there on, we implemented some optional constraints that can help the PTO in finding an optimal solution for their use case, the first one being a constraint that allows a fixed number of non-electric vehicles in Equation 6. Since those non-electric vehicles have no range limitation that is relevant to the PTO's regular operations, they do not need chargers on their routes and usually have slightly lower costs than the electric buses. We also introduced constraints for placing a charger (Equation 7) at a specific location, which might be useful if that charger already exists in the city. Another constraint prohibits specific bus stops from being used for charging (Equation 8). In urban areas it is common that only a few selected bus stops qualify for a charging station, because of spatial limitation, electrical limitations, or opposition from the citizens.

Solution Interpretation

This section describes the output of the ILP model as well as the enrichment and processing for easier interpretation. The model outputs all values for tensors A, B, and C, which hold all the optimization variables. All the values in A hold the number of chargers and the respective type of charger needed for the found optima. B holds the respective vehicles needed for serving the public transport needs. C is a 4-dimensional sparse tensor which holds all values of when and which charger is used in the found solution, thus it is possible to calculate the energy demand at each node. In a helper variable, we stored the value of the max usage of the battery in percent and of the battery capacity that is left at the end of the day, to estimate overnight charging demand.

These optimization variables hold all the information needed to set up the new bus and charger configuration. In our scenario there are more than 30,000 variables, but this number heavily depends on the number of input vehicle types and charger types. Even if only the variables that are used in the solution are considered, it can be hard to grasp the solution. Therefore, we created different visualizations. In Figure 3, a 3D bar chart shows all the possible charger locations as points on the map and all chargers with their usage in minutes over the course of one day. This makes it easier to compare different solutions and maybe restrict the algorithm to a subset of the charger locations to get a more stable solution output.

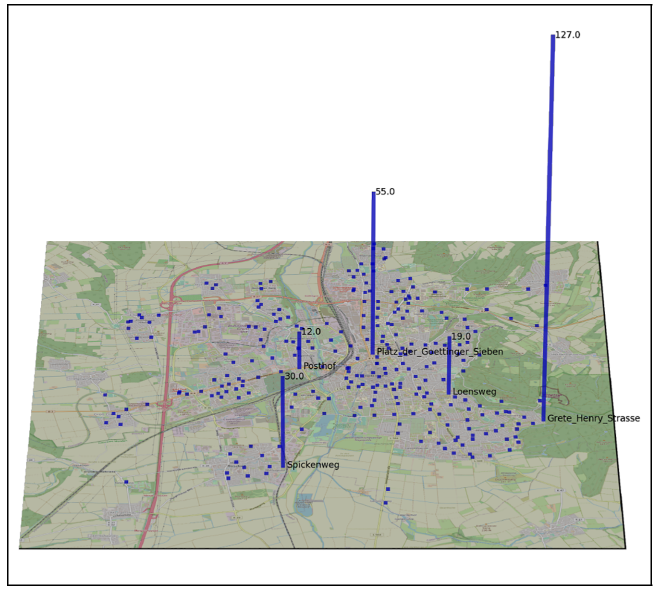

Figure 3. Visualization of charger usage for one solution. A bar represents a charger, the squares represent bus stops and therefore potential charger locations. The time of charging at this bus stop, during a whole day (in minutes) over all vehicles, is visualized by the height of the bar.

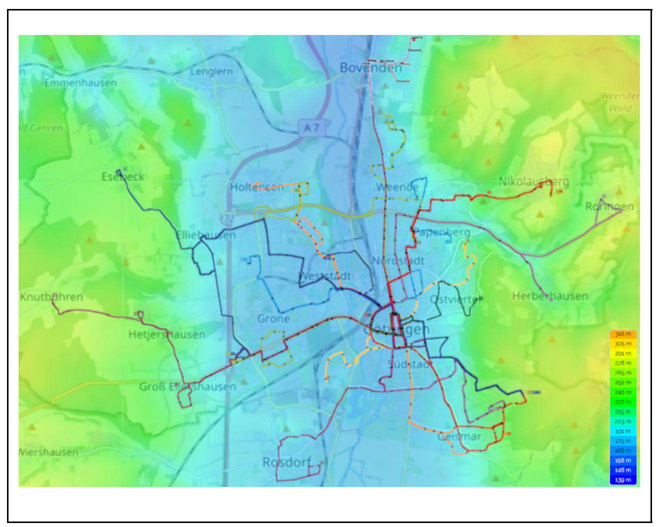

Figure 4. Topology of Göttingen with a public transport bus lines overlay. Almost all lines have substantial differences in altitude.
Source: Graphic was created from excerpts of de-de.topographic-map.com/ and the GöVB bus lines.

Case Study

To understand the scope of the project better, we will briefly present the studied city and some boundary conditions. This study is carried out as accompanying research of a national project to help cities and public transport providers to make the best decisions while switching to battery-electric vehicles. The studied area is the city of Göttingen, which, as can be seen in Figure 4, has a topology that is strongly varying in altitude. The difference in altitude ranges from 150 m above sea level to almost 300 m in the outer parts of the bus network. The local PTO serves customers on 20 lines during the day and eight lines at night with varying frequency between every 15 to every 60 min. This adds up to a weekly mileage of around 26,000 km. To optimize vehicle usage and working hours of the drivers, each line does not get served by one vehicle all day, as this would cause unnecessary dwell times at each end point of a bus line. Instead, all the lines are combined to create rotation plans which ensure optimal usage of the vehicles while considering all regulations on working hours and "proper" lunch break spots, where drivers can get something to eat and rest before their next shift. These regulations were one of the main reasons why we decided to leave the network and frequency-setting problem (*4*) untouched and concentrate on the optimization of charging infrastructure and vehicle choice.

Besides reducing complexity, this should also give more applicable solutions in the real world, since the PTO can check their current plan data without any changes. If they decide to optimize the routes for electric vehicles, those newly created plans can serve as a basis for further optimization steps. In Göttingen, there are 125 rotation plans defined to serve the public transport demand. The length of each rotation differs from 50 to 353 kilometers per day. Although there are buses on the market able to serve all these without recharging, it is not evident if employing these would make for the best scenario for the PTO because battery size is still the most expensive part of battery-electric buses. With the wide variety of possible chargers and buses, it is clear that careful planning is needed to get the best possible outcome. The PTO in our project currently has 90 combustion engine vehicles and 10 hybrid vehicles in use. We equipped 50 of those vehicles with data loggers which track all necessary data and sent it via mobile communication to our servers. Proper coverage of all rotation plans could be ensured without collecting data from every bus, since the vehicles serve different rotations every day.

Results

The main goal of this section is to showcase the usage and flexibility of our framework. We constructed a scenario where the goal was to analyze the impact of stepwise electrification with different operational strategies in mind. Afterward, we evaluated the robustness and computation time for different inputs, to judge the applicability to other real-world scenarios including larger bus networks.

Gradual Electrification

In this scenario we wanted to analyze the effect of an allowed number of diesel buses on the general solution. The experiment is divided into three categories.

1. Mixed Fleet (MF): The algorithm searches for an optimal solution with all available vehicles and chargers defined in the input data set.
2. Depot Charger (DC): Which is commonly defined as an operating strategy where the vehicles do not charge during the day, but only overnight at the depot. This implies low charging powers and high battery capacity. For the implementation of this category, we created three different scenarios where we provide only one charger type with a charging power of 150 kW. For the three scenarios, we varied the battery capacity of the vehicles in a range that seemed realistic for applications at the time (2021): 350, 380, and 410 kWh.
3. Opportunity Charging (OC): In contrast to the DC strategies, OC strategies rely on high charging powers on the route, to reduce the required battery capacity in the vehicles. For this category, we also made three scenarios but this time with a fixed battery capacity of 170 kWh and varied the charging power of the chargers between 300 kW (implemented in our project), 450 kW, and 600 kW. At the time of investigation, these were the chargers with the highest available charging power.

We computed results for all these scenarios with the number of allowed diesel buses changing gradually from 0 to 10 to see how this affected the overall configuration of the algorithm. The results of this experiment are summarized in Figure 5.

As expected, the first category gives the most cost-effective solutions, as it can choose from all vehicle types and charger types. Some interesting findings are that with the given vehicles and routes, a complete DC scenario electrification of the fleet is not possible. Even with the biggest battery size, at least two diesel buses are required to avoid any on-route charging and therefore enable pure DC strategies. Although the buses with small batteries are considerably cheaper than the ones from the DC scenarios, the OC strategies are more expensive without diesel bus support but benefit significantly from a few diesel buses. Plus, at around four to six allowed diesel buses, the OC strategies become cheaper because they can also reduce the number of required chargers.

Robustness

A key factor of the algorithm is the reliability of the solutions. This is usually examined by a robustness analysis. The idea behind this is to find out if only small changes of input parameters would render the found solution infeasible. This is very unlikely as the way the problem is set up still allows enough safety margins for extreme

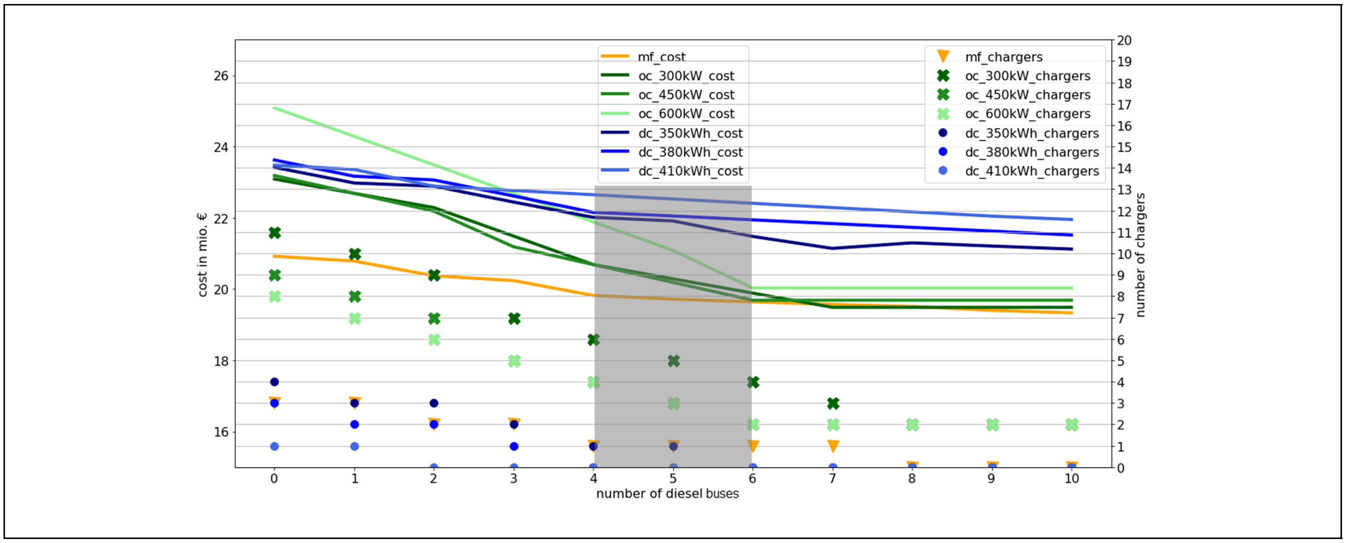

Figure 5. Comparison of total cost of fleet and infrastructure with different operational strategies and different amounts of allowed diesel buses. The *x*-axis shows the number of diesel buses used in the optimization. The first category is shown in yellow where the lines belong to the left *y*-axis and show the cost of the solution, while the different markers in the same color (belonging to the right *y*-axis) denote the number of chargers required in that result. Green shows the OC scenarios, whereas blue represents the DC scenarios.
Note: mf = mixed fleet; oc = opportunity charging; dc = depot charger.

cases. Since the battery is only used between 20% and 80% SOC, it would be possible to drive further if necessary. However, to preserve the longevity of the battery, this is not recommended. Taking this into consideration, we decided against a classical robustness analysis and instead took a closer look at the different solutions and how they influence the actual implementation. The most important factor in this is the locations of the chargers, as this should be a more long-term investment compared to the battery sizes in the vehicles. In our described experiment, we found out that for different levels of electrification (number of allowed diesel buses) the algorithm can suggest different charging locations, which is detrimental to the adoption by a PTO. Therefore, we implemented a view in which we can visualize the chargers and their usage on a map to analyze the solutions as in Figure 3. Then manual restrictions of the options can be applied, and the optimization starts over. This can result in very stable solutions with minimal change in the overall cost. To have a more seamless workflow within the planning of a gradual electrification of the bus network we added a workflow in which the algorithm starts with a high number of allowed diesel buses and saves the charging infrastructure. Then the number of allowed diesel buses gets decreased, and the previous charging infrastructure gets fixed with the constraint in Equation 7. Therefore, the results for the gradual electrification are incremental and do not vary in the placement of chargers.

Computation Time

This section gives an overview of the applicability to larger networks according to computation time. Since this is not a process that needs to be calculated often, a long computation time is acceptable within reason. The experiments were conducted on a server virtual machine with limited resources (4 CPU cores, 16 GB RAM) and can, therefore, give closely comparable results between the different runs. We found the number of input vehicles and chargers to have great impact on the performance, especially when the different performance or cost values are close together. This supposedly stems from the inner workings of the solver. It uses a branch-and-bound approach which tries to eliminate bad solutions early on. This becomes harder with the rising number of similar input values and can be observed in the relation between computation time and the number of branch-and-bound nodes. We observed that the ILP model can, with the whole bus network, actually be solved to optimality in just a few seconds when the number of bus types and charger types are low. The branching algorithm can also eliminate very expensive vehicles or chargers early on when they cannot provide the optimal solution. But if we introduce many different input parameters with minimal difference between the values, the algorithm starts to build up a large decision tree and the computation can take several days (number of vehicle types >10, number of charger types >10). We were able to partially circumvent this by starting off with four to five vehicle types and the same amount of charger types. After an optimal solution was found, we changed the input vehicles and chargers to values closer to the optimal amount and started over. This could potentially also be done by implementing a custom branching logic within the SCIP solver, which we would like to investigate in future work.

Conclusion

This paper presents a general framework for optimizing the transition of an entire existing urban bus network to electric vehicles. In previous studies using LP models, only parts of an urban bus network were analyzed; we could show that this approach also works for a complete, medium-sized city. Large amounts of data were leveraged to get a better understanding of the whole system before trying to find an optimal solution. The big-data approach is a great advantage for this project, since it gives us the opportunity to validate many assumptions about the detailed energy demand, the timeliness of the buses, and therefore potential charging times and other operational details. This gives more reliable results than standardized values for energy consumption like the standardized on-road test cacles (SORT) driving cycles (*37*), especially for areas with high deviation in altitude. The experiments conducted on the obtained data show that for a medium-sized German city we can produce optimal solution with low computation time. We could demonstrate that this approach can be incorporated in the planning process of most PTOs, because the algorithm starts off with an existing bus network and allows for flexible integration of constraints like feasible charger locations or the number of acceptable diesel buses.

Currently, we are working on further data analyses of the bus data, to include more detailed energy demands for different times of the day, since it would be beneficial for the planners to have a concise overview of the operational conditions of the urban bus network. Aside from that, we are currently developing a traffic simulation environment based on Simlation of Urban MObility (SUMO), which should enable the verification of our algorithm for new bus routes, where we do not have any tracking data. This simulation incorporates a detailed traffic model, a street network, and the actual buses that drive according to the rotation plans.

In the future, the model will be extended to evaluate different technologies like inductive charging at bus stops, traffic lights, or during driving. This could enable completely new operating strategies. We also plan on tweaking the optimizer itself, in particular the branching rules to be able to find optimal solutions with greater number of input parameters in a reasonable amount of time.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: S. Würtz, K. Bogenberger, U. Göhner; data collection: S. Würtz; analysis and interpretation of results: S. Würtz; draft manuscript preparation: S. Würtz. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was conducted in the context of the project OmniE (www.omnie.de) and funded by the German Federal Ministry for Economic Affairs and Energy. https://www.digitale-technologien.de/DT/Redaktion/DE/Standardartikel/IKT-EM/IKT-Projekte/ikt-III-projekt-OmniE.html

ORCID iDs

Samuel Würtz https://orcid.org/0000-0002-6610-9100
Klaus Bogenberger https://orcid.org/0000-0003-3868-9571
Ulrich Göhner https://orcid.org/0000-0002-4795-4798

References

1. The European Commission. *Clean Vehicle Directive*. EC, 2019. https://ec.europa.eu/transport/themes/urban/clean-vehicles-directive_en.
2. Gorguinpour, C., S. Castellanos, X. Li, and R. Sclar. *How to Enable Electric Bus Adoption in Cities Worldwide: A Guiding Report for City Transit Agencies and Bus Operating Entities*. World Resources Institute, Washington, D.C., 2019.
3. Tenzer, F. *Prognose zum weltweit generierten Datenvolumen 2025*. Statista, 2018. https://de.statista.com/statistik/daten/studie/267974/umfrage/prognose-zum-weltweit-\generierten-datenvolumen/.
4. Ceder, A., and N. H. Wilson. Bus Network Design. *Transportation Research Part B: Methodological*, Vol. 20, No. 4, 1986, pp. 331–344. https://doi.org/10.1016/0191-2615(86)90047-0.
5. Mandl, C. E. Evaluation and Optimization of Urban Public Transportation Networks. *European Journal of Operational Research*, Vol. 5, No. 6, 1980, pp. 396–404. https://doi.org/10.1016/0377-2217(80)90126-5.
6. Iliopoulou, C., I. Tassopoulos, K. Kepaptsoglou, and G. Beligiannis. Electric Transit Route Network Design Problem: Model and Application. *Transportation Research Record: Journal of the Transportation Research Board*, 2019. 2673: 264–274.
7. Ahmed, L., P. Heyken-Soares, C. Mumford, and Y. Mao. Optimising Bus Routes With Fixed Terminal Nodes. In

Proceedings of the Genetic and Evolutionary Computation Conference (M. López-Ibáñez, A. Auger, and T. Stützle, eds.), ACM, New York, NY, 2019, pp. 1102–1110. https://doi.org/10.1145/3321707.3321867.

8. Uslu, T., and O. Kaya. Location and Capacity Decisions for Electric Bus Charging Stations Considering Waiting Times. *Transportation Research Part D: Transport and Environment*, Vol. 90, 2021, P. 102645. https://doi.org/10.1016/j.trd.2020.102645. http://www.sciencedirect.com/science/article/pii/S1361920920308300.
9. Kameda, H., and N. Mukai. Optimization of Charging Station Placement by Using Taxi Probe Data for On-Demand Electrical Bus System. In *Proc., Lecture Notes in Computer Science Lecture Notes in Artificial Intelligence: International Conference on Knowledge-Based and Intelligent Information and Engineering Systems* (A. König, ed.), Kaiserslautern, Germany, September 12–14, 2011, Springer, Berlin, Heidelberg, pp. 606–615.
10. Sadeghi-Barzani, P., A. Rajabi-Ghahnavieh, and H. Kazemi-Karegar. Optimal Fast Charging Station Placing and Sizing. *Applied Energy*, Vol. 125, 2014, pp. 289–299. https://doi.org/10.1016/j.apenergy.2014.03.077.
11. Kunith, A., R. Mendelevitch, A. Kuschmierz, and D. Göhlich. Optimization of Fast Charging Infrastructure for Electric Bus Transportation – Electrification of a City Bus Network. *Proc., EVS29 International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium*, Montréal, Québec, Canada, 2016.
12. Gao, Z., Z. Lin, T. J. LaClair, C. Liu, J.-M. Li, A. K. Birky, and J. Ward. Battery Capacity and Recharging Needs for Electric Buses in City Transit Service. *Energy*, Vol. 122, 2017, pp. 588–600. https://doi.org/10.1016/j.energy.2017.01.101.
13. de Filippo, G., V. Marano, and R. Sioshansi. Simulation of an Electric Transportation System at The Ohio State University. *Applied Energy*, Vol. 113, 2014, pp. 1686–1691. https://doi.org/10.1016/j.apenergy.2013.09.011.
14. Jefferies, D., and D. Göhlich. A Comprehensive TCO Evaluation Method for Electric Bus Systems Based on Discrete-Event Simulation Including Bus Scheduling and Charging Infrastructure Optimisation. *World Electric Vehicle Journal*, Vol. 11, No. 3, 2020, P. 56. https://doi.org/10.3390/wevj11030056.
15. Berthold, K. *Techno-ökonomische Auslegungsmethodik für diëElektrifizierung urbaner Busnetze, Karlsruher Schriftenreihe Fahrzeugsystemtechnik*, Vol. 74. KIT Scientific Publishing, Karlsruhe, Baden, 2019.
16. Kunith, A. W. *Elektrifizierung des urbanen öffentlichen Busverkehrs*. Dissertation. Springer Fachmedien Wiesbaden GmbH, Germany, 2017.
17. Rogge, M. *Elektrifizierung von ÖPNV-Busflotten mit batterieelektrischen Bussen -Entwicklung einer Software-Toolchain für die Umstellungsplanung ganzer Busflotten unter Berücksichtigung von technischen und betrieblichen Randbedingungen: Institut für Stromrichtertechnik und f Elektrische Antriebe / Lehrstuhl für Elektrochemische Energiewandlung und Speichersystemtechnik*. PhD thesis, ISEA and dissertation. Rheinisch-Westfälische Technische Hochschu Aachen, Germany, 2020. https://publications.rwth-aachen.de/record/814282/export/hx.
18. Kunith, A., R. Mendelevitch, and D. Goehlich. Electrification of a City Bus Network—An Optimization Model for Cost-Effective Placing of Charging Infrastructure and Battery Sizing of Fast-Charging Electric Bus Systems. *International Journal of Sustainable Transportation*, Vol. 11, No. 10, 2017, pp. 707–720. https://doi.org/10.1080/15568318.2017.1310962.
19. Kessler, L., and K. Bogenberger. Dynamic Traffic Information for Electric Vehicles as a Basis for Energy-Efficient Routing. *Transportation Research Procedia*, Vol. 37, 2019, pp. 457–464. https://doi.org/10.1016/j.trpro.2018.12.218.
20. Morlock, F., B. Rolle, M. Bauer, and O. Sawodny. Forecasts of Electric Vehicle Energy Consumption Based on Characteristic Speed Profiles and Real-Time Traffic Data. *IEEE Transactions on Vehicular Technology*, Vol. 69, No. 2, 2020, pp. 1404–1418. https://doi.org/10.1109/TVT.2019.2957536.
21. NASA JPL. *NASA Shuttle Radar Topography Mission Global 1 arc Second*. SRTMGL1, 2013. https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003. https://cmr.earthdata.nasa.gov/search/concepts/C1220567890-USGS_LTA.html.
22. OpenStreetMap. Map Dump. 2021. https://planet.osm.org.
23. Apache-Softwarefoundation. Apache Kafka. 2021. https://kafka.apache.org/.
24. Timescale-Inc. *PostrgeSQL*. Timescale, 2021. https://www.timescale.com/.
25. HDEI-BCEI-Task-Force. Bus-FMS-Standard. 2021. http://www.fms-standard.com/Bus/index.htm.
26. VDV -Verband Deutscher Verkehrsunternehmen. *VDV-Schnittstelleninitiative: Soll-Daten-Schnittstellen*. VDV, 2021. https://www.vdv.de/oepnv-datenmodell.aspx.
27. MATLAB. *Version 7.10.0 (R2010a)*. The MathWorks Inc, Natick, MA, 2010.
28. VDV -Verband Deutscher Verkehrsunternehmen. *Klimatisierung von Linienbussen der Zulassungsklassen I (Stadtbus) und II (Überlandbus, für konventionell angetriebene Dieselund Gasbusse als auch für Hybrid-, Brennstoffzellen und Elektrobusse*. VDV-Schrift 236, 11/2018. VDV, Köln, Nordrhein-Westfalen, 2018.
29. van Rossum, G., and F. L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.
30. Perron, L., and V. Furnon. OR-Tools. 2019. https://developers.google.com/optimization/. https://developers.google.com/optimization/.
31. SCIP. *Solving Constraint Integer Programs*. Zuse Institute Berlin (ZIB), July, 2021. https://scipopt.org/.
32. Gamrath, G., D. Anderson, K. Bestuzheva, W.-K. Chen, L. Eifler, M. Gasse, P. Gemander, et al. *The SCIP Optimization Suite 7.0*. Technical Report. Mathematical Optimization Society, Philadelphia, PA, 2020. http://www.optimization-online.org/DB_HTML/2020/03/7705.html.
33. Kanapady, R., K. Y. Kyle, and J. Lee. Battery Life Estimation Model and Analysis for Electronic Buses With Auxiliary Energy Storage Systems. *Proc., 2017 IEEE Applied Power Electronics Conference and Exposition (APEC)*,

Tampa, FL, IEEE, New York, 2017, pp. 945–950. https://doi.org/10.1109/APEC.2017.7930810.

34. Tang, L., G. Rizzoni, and A. Cordoba-Arenas. Battery Life Extending Charging Strategy for Plug-In Hybrid Electric Vehicles and Battery Electric Vehicles. *IFAC-PapersOnLine*, Vol. 49, No. 11, 2016, pp. 70–76. https://doi.org/10.1016/j.ifacol.2016.08.011.
35. Bloom, I., B. Cole, J. Sohn, S. Jones, E. Polzin, V. Battaglia, G. Henriksen, et al. An Accelerated Calendar and Cycle Life Study of Li-Ion Cells. *Journal of Power Sources*, Vol. 101, No. 2, 2001, pp. 238–247. https://doi.org/10.1016/S0378-7753(01)00783-2.
36. Fasthuber, D., and M. Litzlbauer. *Erkenntnisse der Messung von Ladevorgaengen der Elektrofahrzeuge in der Modellregion e-pendler in niederoesterreich*. PhD thesis. TU Wien, 2016. https://www.tugraz.at/fileadmin/user_upload/Events/Eninnov2016/files/pr/Stream_G/Session_G2/PR_Fasthuber.PDF.
37. UITP. Standardised On-Road Test Cycles-SORT: A Project of the UITP Bus Committee in Collaboration With Manufacturers. 2001. https://ec.europa.eu/environment/archives/clean_bus/slides/etienne_sort.pdf.