Designing a Robust Charging Infrastructure for Battery-Eelectric Buses with Sparse Energy Consumption Data

Sara Momen*[1], Yousef Maknoon[2], Bart van Arem[1], and Shadi Sharif Azadeh[1]

[1]Transport and Planning Department, Delft University of Technology, Netherlands
[2]Engineering, Systems and Services Department, Delft University of Technology, Netherlands

Short summary

This paper tackles the challenges of designing charging infrastructure for Battery Electric Buses (BEBs) with limited energy consumption data. Accurate energy consumption estimation is critical for cost-effective and reliable electrification but often requires costly field experiments, leading to limited data. To address uncertainty and data sparsity, we propose two models: a robust optimization model with box uncertainty, which relies solely on the data range, and a data-driven distributionally robust optimization model that uses observed data for more informed solutions. Our analysis of the Rotterdam bus network reveals three key insights: (1) Ignoring energy consumption variations can lead to unreliable designs. (2) Designing infrastructure based on worst-case energy consumption increases costs by 67% compared to using average estimates. (3) The data-driven distributionally robust optimization model is more cost-effective than the box uncertainty model while maintaining reliability, even when extreme energy consumption values are frequent.
Keywords: Public transport electrification, Energy consumption variability, Operational reliability with sparse data, Distributionally robust optimization

1 Introduction

In the paradigm of zero-emission urban mobility, Battery Electric Buses (BEBs) are promising due to their independence from traditional infrastructure like cables and rails. However, designing effective and reliable BEB networks poses challenges. A key issue is estimating the energy consumption and strategically placing charging stations to ensure continuous service. Accurately estimating energy consumption is critical for the cost-efficient and reliable deployment of electrified transport networks. This estimation is challenging due to a wide range of influential factors, including traffic conditions, passenger load, and weather. Inaccurate estimates of energy consumption impact electrification costs, battery lifetime, and service levels (Azadeh et al., 2022).
Robust optimization, addresses uncertainty by solving problems under the worst-case realization of random variables. This method has been widely applied to BEB-related challenges, with studies incorporating diverse sources of uncertainty into their formulations, as shown in Hu et al. (2022); Bai et al. (2022); Avishan et al. (2023). While robust optimization is computationally efficient and well-suited to data-scarce environments, its reliance on worst-case scenarios often leads to overly conservative solutions that fail to capture the probabilistic nature of real-world uncertainty (Birge & Louveaux, 2011). To address these shortcomings, distributionally robust optimization (DRO) incorporates distributional information to balance robustness with performance, offering a more flexible approach to decision-making under uncertainty.
This study first models the CID problem using average energy consumption estimates, recognizing that ignoring energy consumption uncertainty can lead to infeasible designs. To address this in data-scarce environments, we propose two robust optimization models for the CID problem. The first, a box uncertainty model with a budget (BoU-CID), uses a range-based uncertainty framework, effective when only the range of uncertainty is known but may lead to overly conservative designs. The second, a data-driven distributionally robust chance constraint approach (DRCC-CID), incorporates observed data characteristics, offering more cost-effective and reliable solutions. Both models are evaluated across various scenarios to assess their performance in terms of cost-efficiency and reliability.

2 Methodology

We formulate the CID problem for the electrification of an existing bus network, addressing the challenges posed by sparse energy consumption data. The problem involves three core decisions: (1) selecting the type of charging station (e.g., fast-feeding or standard), (2) identifying optimal charging station locations, and (3) determining onboard battery capacities.

Formulation using average energy consumption estimates

This model utilizes expected values to estimate uncertain energy consumption, serving as a baseline to gauge operational efficiencies. The objective function of all the models (1) aims to minimize the total cost, compromising the installation cost of type $t \in T$ of charging stations (denoted by α_t) and onboard battery costs per kWh (denoted by β). The BEB fleet size for each line $k \in K$ is given and denoted by γ_k. The binary variable x_{st} is set to 1 if a charging station type $t \in T$ is installed at stop s. The battery capacity for each bus line is represented by the continuous variable z_k.

$$\min_{x,z} \quad \sum_{s \in S} \sum_{t \in T} \alpha_t x_{st} + \sum_{k \in K} \beta \gamma_k z_k \tag{1}$$

Charging station. We assume that at most one type of charging station can be installed at each stop. This condition is enforced by Constraint 2.

$$\sum_{t \in T} x_{st} \leq 1, \qquad \forall s \in S \tag{2}$$

$$x_{st} \in \{0, 1\}, \qquad \forall s \in S, \forall t \in T \tag{3}$$

Energy flow. The variable e_{ks} represents the energy level of the BEB on line k as it departs stop s. Constraint (4) governs the energy flow between stops $s-1$ and s. It ensures that the energy level at departure from stop s is the energy level at departure from stop $s-1$ minus the expected energy consumption between these stops, denoted by $\bar{\mu}_s^k$, plus any energy gained at stop s. The power of the charging station type t is denoted by P_t, and Δ_{ks} indicates the dwelling time at stop s for bus line k. For fast-feeding charging stations, the BEB departs with a fully charged battery regardless of the dwelling time, which is modeled as $P_{FF} = \bar{b} z_k$. BEBs must depart the terminal with maximum energy levels, given by $e_{k o_k} = \bar{b} z_k, \quad \forall k \in K$.
Using the energy flow Constraint (4), e_{ks} is computed cumulatively from the terminal, as defined in Constraint (5).

$$e_{ks} \leq e_{ks-1} - \bar{\mu}_s^k + \sum_{t \in T} P_t \Delta_{ks} x_{st}, \qquad \forall k \in K, s \in S_k \tag{4}$$

$$e_{ks} = \bar{b} z_k - \sum_{i=o_k}^{s} \bar{\mu}_i^k + \sum_{t \in T} \sum_{i=o_k}^{s} P_t \Delta_{ki} x_{it}, \qquad \forall k \in K, s \in S_k \tag{5}$$

Battery energy level. The BEB energy level at all stops must not exceed the maximum allowed, expressed as $e_{ks} \leq \bar{b} z_k, \quad \forall k \in K, s \in S_k \setminus o_k$. When applied to Constraint 5, the term $\bar{b} z_k$ cancels out, simplifying the constraint. This ensures that the cumulative energy gained through charging up to stop s does not exceed the cumulative average energy consumption up to stop s, thereby preventing overcharging and adhering to the maximum allowed battery capacity, as shown in Constraint (6). Constraint (7) ensures that the energy level upon arrival at any stop s remains above the minimum required.

$$\sum_{t \in T} \sum_{i=o_k}^{s} P_t \Delta_{ki} x_{it} \leq \sum_{i=o_k}^{s} \bar{\mu}_i^k \qquad \forall k \in K, s \in S_k \tag{6}$$

$$\bar{b} z_k - \sum_{i=o_k}^{s} \bar{\mu}_i^k + \sum_{t \in T} \sum_{i=o_k}^{s-1} P_t \Delta_{ki} x_{it} \geq \underline{b}\, z_k \qquad \forall k \in K, s \in S_k \tag{7}$$

$$z_k \in \mathbb{R}^+, \qquad \forall k \in K \tag{8}$$

The (deterministic) CID is to minimize 1 subject to (2, 3) and (6-8) set of constraints.

Formulation based on uncertainty range

This strategy is applicable when the range of energy consumption data is known to decision-makers. For two subsequent stops $s-1$ and s, the uncertainty box is defined by the expected energy consumption ($\bar{\mu}_s^k$) and its maximum deviation ($\hat{\mu}_s^k$). To limit the deviation, we impose Γ_s^k as a budget to set an upper limit for energy consumption deviations up to stop s on bus line k. We define the box uncertainty set for φ_n^k as $\mathcal{C}(\varphi_n^k)$:

$$\mathcal{C}(\varphi_n^k) = \{\varphi_i^k \in [0,1] | \sum_{i=o_k}^{s} \varphi_i^k \leq \Gamma_s^k, \quad \forall k \in K\}$$

The parameter Γ_s^k belongs to the interval $[0, |s|]$. This parameter indicates the cumulative allowed energy consumption deviation up to stop s. The value of Γ_s^k depends on decision-maker opinion and indicates the robustness and level of conservatism of the solution. The range of the energy consumption random parameter needs to be considered in (6) and (7) using the defined box uncertainty set $\mathcal{C}(\varphi_n^k)$. The final BoU-CID formulation is as follows, where u^k and v_i^k are the dual variables.

$$BoU - CID = 1$$

subject to: (2, 3, 8)

$$\sum_{t\in T}\sum_{i=o_k}^{s} P_t \Delta_{ki} x_{it} \leq \sum_{i=o_k}^{s} \bar{\mu}_i^k + \Gamma_s^k u^k + \sum_{i=o_k}^{s} v_i^k, \qquad \forall k \in K, s \in S_k \tag{9}$$

$$(\bar{b} - \underline{b}) z_k + \sum_{t\in T}\sum_{i=o_k}^{s-1} P_t \Delta_{ki} x_{it} \geq \sum_{i=o_k}^{s} \bar{\mu}_i^k + \Gamma_s^k u^k + \sum_{i=o_k}^{s} v_i^k, \quad \forall k \in K, s \in S_k \tag{10}$$

$$u^k + v_s^k \geq \hat{\mu}_s^k, \qquad \forall k \in K, s \in S_k \tag{11}$$

$$u^k, v_s^k \geq 0, \qquad \forall s \in S_k \tag{12}$$

Empirical data-driven uncertainty modeling

In the previous formulation, we demonstrated the importance of accounting for values above the average consumption to ensure robustness of the design and developed a tractable formulation based on this concept. However, this model may be vulnerable to inaccuracies in estimating the worst-case scenarios. Unlike stochastic programming, which relies on predefined probability distributions, and robust optimization, which assumes worst-case parameter realizations, DRO reduces conservatism while maintaining computational tractability. This makes it particularly effective for problems with limited or uncertain data (Delage & Ye, 2010)

We consider N samples, each representing observations of energy consumption between consecutive bus stops on a specific bus line. The j-th energy consumption sample between stop $s-1$ and s on bus line k is noted by μ_s^{kj}, where $\{j = 1, ..., N\}$. It ensures that the cumulative energy consumption up to stop s on line k falls within a decision-dependent safety set with high probability $(1-\epsilon)$ across all potential distribution (P), as shown in Constraints (13) and (14). In these constraints, the average energy consumption in **CID** is replaced by the observed values of energy consumption. The related risk parameter is shown by ϵ.

$$P\left[\sum_{t\in T}\sum_{i=o_k}^{s} P_t \Delta_{ki} x_{it} \leq \sum_{i=o_k}^{s} \mu_i^{kj}, \quad \forall k \in K, s \in S_k, j \in \{1, ..., N\}\right] \geq 1-\epsilon, \quad \forall P \in \mathcal{F}(\theta_s^k) \tag{13}$$

$$P\left[(\bar{b} - \underline{b}) z_k + \sum_{t\in T}\sum_{i=o_k}^{s-1} P_t \Delta_{ki} x_{it} \geq \sum_{i=o_k}^{s} \mu_i^{kj}, \forall k \in K, s \in S_k, j \in \{1, ..., N\}\right] \geq 1-\epsilon, \quad \forall P \in \mathcal{F}(\theta_s^k) \tag{14}$$

Given the N observations of energy consumption data, accurately estimating their true distribution, which is influenced by multiple factors, is complex. To address this, we define a distributional uncertainty set, or ambiguity set (denoted by $\mathcal{F}(\theta_s^k)$), that includes all possible distributions for energy consumption (P) within a specified distance (θ_s^k) from a reference distribution, denoted by

$\hat{P}$. The reference distribution is modeled as a discrete uniform distribution calculated from the energy consumption observations, expressed as $\hat{P} = \frac{1}{N}\sum_{j=1}^{N}\delta_{\mu_s^{kj}}$, where δ represents the Dirac delta function for each energy consumption observation μ_s^{kj}. The ambiguity set is defined by $\mathcal{F}(\theta_s^k) = \{P : d_W(\hat{P}, P) \leq \theta_s^k\}$, with d_W being the Wasserstein metric measuring distributional distance. This metric effectively forms a ball of radius $\theta_s^k \geq 0$ centered on the empirical distribution $\hat{P}$.

The decision-dependent safety set ensures that constraints (13) and (14) are met. This allows the model to probe the boundary of the safety set and identify the worst probable distribution of energy consumption influenced by the parameter θ_s^k. The parameter θ_s^k serves as a budget for displacement, dictating the extent to which observed energy data up to stop s on bus line k can shift toward the boundary of its safety set. The decision maker determines the value of θ_s^k, which sets the size of the ambiguity set. A larger θ_s^k value provides more budget to transport an energy consumption sample to its worst value, leading to more conservative solutions. We present the final formulations in the following:

$$\epsilon N q^k - \sum_{i=o_k}^{s}\sum_{j\in\{1,\ldots,N\}} r_i^{kj} \geq \theta_s^k N \qquad \forall k \in K, s \in S_k \tag{15}$$

$$M(1 - \sum_{i=o_k}^{s} y_i^{kj}) \geq q^k - \sum_{i=o_k}^{s} r_i^{kj} \qquad \forall k \in K, s \in S_k, j \in \{1, \ldots, N\} \tag{16}$$

$$-\sum_{t\in T}\sum_{i=o_k}^{s} P_t \Delta_{ki}\, x_{it} + \sum_{i=o_k}^{s} \mu_i^{kj} + M \sum_{i=o_k}^{s} y_i^{kj} \geq q^k - \sum_{i=o_k}^{s} r_i^{kj} \qquad \forall k \in K, s \in S_k, j \in \{1, \ldots, N\} \tag{17}$$

$$(\bar{b} - \underline{b})z_k + \sum_{t\in T}\sum_{i=o_k}^{s-1} P_t \Delta_{ki} x_{it} - \sum_{i=o_k}^{s} \mu_i^{kj} + M \sum_{i=o_k}^{s} y_i^{kj} \geq q^k - \sum_{i=o_k}^{s} r_i^{kj} \qquad \forall k \in K, s \in S_k, j \in \{1, \ldots, N\} \tag{18}$$

$$y_s^{kj} \in \{0,1\}^N,\; r_s^{kj},\; q^k \geq 0, \qquad \forall k \in K, s \in S_k \tag{19}$$

where $q^k \geq 0$ and $r_s^{kj} \geq 0$ are the dual variables for the reformulation of the distance to the unsafe set. Following Chen et al. (2022), we introduce a binary variable y_s^{kj}, where $y_s^{kj} = 1$ indicates that the sample μ_s^{kj} does not satisfy the chance constraint, and vice versa. Constraints (15) and (16–19) limit the distance of observations to the unsafe set, ensuring that the probability of not satisfying the chance constraint is smaller than ϵ. Constraints (16–18) include a sufficiently large constant $M \in \mathbb{R}^+$. The $DRCC - CID$ model, is ultimately reformulated as an MILP and presented below:

$$\begin{aligned} DRCC - CID = & 1 \\ & \text{subject to: } (2,\ 3,\ 8), \\ & \qquad (15 - 19) \end{aligned}$$

3 Results and discussion

To demonstrate the practical effectiveness of the formulations, they are applied to Rotterdam bus network. We select three distinct bus lines (33, 38, and 40) within the Rotterdam bus network to investigate the deployment of charging infrastructure for BEBs. All lines begin at the common terminal, *Rotterdam Centraal Station*, and may share stops with one another. Each line is serviced by a fleet of 10 buses.

Various methods have been proposed for estimating energy consumption in BEBs. By treating energy consumption as an exogenous input to the model, we reduce dependency on detailed parameter estimation. The average energy consumption ($\bar{\mu}$) is simulated based on the distance between two consecutive stops, with a consumption rate of 1.3 kWh per km (Bai et al., 2022). In the DRCC-CID the base scenario involves generating 100 energy consumption samples from a uniform distribution within the range $[\bar{\mu}, \hat{\mu}]$ for each pair of stops. The risk tolerance ϵ is set to 0.1 to satisfy the chance constraints, and M is fixed at 25 after systematic evaluation.

Electrification cost analysis

The total electrification costs of the three models for different budget parameters are shown in Table 1. The first row of the table presents the deterministic CID model results, serving as a benchmark. Column 3 lists the total costs for each model. Column 4 shows the cost difference of solutions compared to the benchmark. Columns 5 and 6 present the percentage of the total electrification cost attributed to charging station installation and BEB battery capacity, respectively. The results show that BoU-CID leads to designs that are 67% more expensive than the CID model under high conservatism, with increases observed in both charging station installation and battery capacity costs. In contrast, even under the more conservative setting, DRCC-CID incurs substantially lower costs compared to BoU-CID. The cost breakdown shows that DRCC-CID primarily addresses uncertainty by optimizing BEB battery capacity, reducing the need for costly charging station installations, and minimizing overall system costs.

Table 1: Electrification costs breakdown of models

Model	**Uncertainty level**	**Total costs (€)**	**Relative difference (%)**	**Cost type percentage**	
				CS installment	BEB battery
CID	-	8.413×10^5	**benchmark**	41%	59%
BoU-CID	low: ($\Gamma = 0.2$)	1.18×10^6	24%	43%	57%
	high: ($\Gamma = 0.8$)	1.40×10^6	67%	48%	52%
DRCC-CID	low: ($\theta = 0.2$)	9.81×10^5	17%	27%	73%
	high: ($\theta = 0.8$)	9.98×10^5	19%	26%	74%

Feasibility analysis of the design

This section explores the performance of the models under unobserved conditions by altering the probability distributions used for sampling. In this modified simulation experiment, the sampling distributions for testing differ from the distribution utilized in designing the models, differing in their supports and distributional characteristics. Table 2 summarizes the findings, with the second column indicating the distribution used to achieve the design decisions of the model and the third column is the sampling distribution used for testing the performance of the model in previously unconsidered scenarios. The sixth column shows the relative difference in total electrification costs for robust models compared to the CID solution. The average feasibility performance across entire bus network is indicated in the last column. The first simulation examine the case where the test energy consumption values can exceed the design data by 20% in the range. The second simulation explores scenario where the test and design distribution differ in their characteristics: a symmetric triangular distribution with a shifted mode.

Table 2: Out-of-sample simulation results for CID, BoU-CID, and DRCC-CID

nr.	**Design distribution**	**Out-of-sampling distribution**	**Uncertainty level of solved problem**		**Total cost**	**Average feasibility of the bus network**
			model		Relative difference	
1	$\mathcal{U}[0, \hat{\mu}]$	$\mathcal{U}[0, \mathbf{1.2}\hat{\mu}]$	**CID**	-	bench.	14%
			BoU-CID	low: ($\Gamma = 0.2$)	+8%	56%
				high: ($\Gamma = 0.8$)	+62%	Fully feasible
			DRCC-CID	low: ($\theta = 0.2$)	+0.2%	37%
				high: ($\theta = 0.8$)	+3%	56%
2	$\mathcal{T}[0, \hat{\mu}, \frac{\hat{\mu}}{2}]$	$\mathcal{T}[0, \hat{\mu}, \mathbf{0.8}\hat{\mu}]$	**CID**	-	bench.	17%
			BoU-CID	low: ($\Gamma = 0.2$)	+27%	89%
				high: ($\Gamma = 0.8$)	+75%	Fully feasible
			DRCC-CID	low: ($\theta = 0.2$)	+15%	40%
				high: ($\theta = 0.8$)	+16%	71%

The DRCC-CID model consistently outperforms the deterministic CID model in all simulations, demonstrating improved feasibility under uncertainty. When test data differ only in range but retain similar distributional properties (simulation number 1), DRCC-CID maintains a reasonable level of feasibility while avoiding the excessive costs associated with BoU-CID.

4 Conclusions

This study highlights the substantial influence of uncertainty modeling approaches on BEB infrastructure design and electrification costs. Future research should explore additional dimensions of uncertainty, such as supply-side factors, and extend to tactical and operational decision-making, including the integration of bus schedules with BEB fleet sizing. While this study assumes a complete transition to BEBs, future work could examine the implications of partial electrification and the associated infrastructure adjustments.

References

Avishan, F., Yanıkoğlu, İ., & Alwesabi, Y. (2023). Electric bus fleet scheduling under travel time and energy consumption uncertainty. *Transportation Research Part C: Emerging Technologies*, *156*, 104357.

Azadeh, S. S., Vester, J., & Maknoon, M. (2022). Electrification of a bus system with fast charging stations: Impact of battery degradation on design decisions. *Transportation Research Part C: Emerging Technologies*, *142*, 103807.

Bai, Z., Yang, L., Fu, C., Liu, Z., He, Z., & Zhu, N. (2022). A robust approach to integrated wireless charging infrastructure design and bus fleet size optimization. *Computers & Industrial Engineering*, *168*, 108046.

Birge, J. R., & Louveaux, F. (2011). *Introduction to stochastic programming.* Springer Science & Business Media.

Chen, Z., Kuhn, D., & Wiesemann, W. (2022). Data-driven chance constrained programs over wasserstein balls. *Operations Research*.

Delage, E., & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, *58*(3), 595–612.

Hu, H., Du, B., Liu, W., & Perez, P. (2022). A joint optimisation model for charger locating and electric bus charging scheduling considering opportunity fast charging and uncertainties. *Transportation Research Part C: Emerging Technologies*, *141*, 103732.