



**HAL**  
open science

# Locating electric vehicle charging stations under uncertain battery energy status and power consumption

Mouna Kchaou Boujelben, Céline Gicquel

► **To cite this version:**

Mouna Kchaou Boujelben, Céline Gicquel. Locating electric vehicle charging stations under uncertain battery energy status and power consumption. *Computers & Industrial Engineering*, 2020, 149, pp.106752. 10.1016/j.cie.2020.106752 . hal-02993211

**HAL Id: hal-02993211**

**<https://hal.science/hal-02993211>**

Submitted on 6 Nov 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Locating electric vehicle charging stations under uncertain battery energy status and power consumption

Mouna Kchaou-Boujelben<sup>1</sup>, Céline Gicquel<sup>2</sup>

<sup>1</sup> United Arab Emirates University, Al Ain, United Arab Emirates

<sup>2</sup> Université Paris Saclay, Orsay, France

**Abstract** Fostering the adoption of electric vehicles (EVs) by private drivers requires the development of a wide network of fast charging stations in which drivers traveling long-distance trips will be able to easily recharge their battery within a few minutes. However, due to their high installation costs, the number of stations that can actually be deployed with the available investment budget is strongly limited. It is thus necessary to carefully choose their location so that the charging demand satisfaction is maximized. This leads to the formulation of a facility location problem known as the flow refueling location problem (FRLP). In this paper, we study the FRLP and seek to take into account uncertainties on the vehicle driving range in the problem modeling. We propose to relax several modeling assumptions previously used in the literature to handle this problem. First, we allow the power consumption on a road segment to depend on the crossing direction. Second, we take into account uncertainties related to the energy available in the battery after recharging at a station as well as uncertainties related to the power consumption on each portion of the road network. Finally, we consider statistical dependencies between the stochastic power consumption on different arcs of the network.

We focus on the chance-constrained flow refueling location model, which seeks to maximize the number of drivers for whom the probability of running out of fuel when carrying out their trip is below a certain threshold. To solve the resulting stochastic optimization problem, we propose to use a solution approach based on a partial sample approximation of the stochastic parameters and compare its performance with the one of a previously published approach based on Bonferroni's inequality. We carry out numerical experiments on a set of medium-size randomly generated and real life instances. Our results show that the proposed partial sample approximation approach outperforms the Bonferroni approach in terms of solution quality and gives station locations which provide a significantly improved demand coverage in practice.

**Keywords:** Flow refueling location model, electric vehicle charging station network design, stochastic driving range, stochastic power consumption, chance-constrained programming, partial sample approximation

## 1 Introduction

In order to reduce CO<sub>2</sub> emissions from transport modes, governments have been encouraging drivers to use Electric Vehicles (EVs) instead of conventional gasoline-powered vehicles. A battery EV is a vehicle whose electric motor is powered by the electricity stored in a battery and which requires to be plugged to an external source of electric power to recharge it. Two main modes of use of EVs have been identified. The first mode corresponds to daily short-distance trips, e.g. to commute between home and work. In this case, the battery can be recharged at home and/or at the workplace using a low or medium-speed charging station. This type of charging stations is thus suitable for residential areas, workplaces and car parks where cars stay for several hours. The present work focuses on a second mode of use which corresponds to long-distance trips. In this case, the battery capacity will not be sufficient to provide all the power needed to carry out the trip and the driver will have to stop multiple times during the trip to recharge his battery. Hence, in order to make EVs an attractive option for long-distance trips, an adequate infrastructure based on stations capable of charging a vehicle within a few minutes needs to be deployed. However, these fast charging stations require high installation and operating costs. It is thus important to carefully select their location so as to simultaneously enable as many EV drivers as possible to carry out their trip and comply with the restrictions on the investment budget. This combinatorial optimization problem, known as the flow refueling location problem (FRLP), was first introduced in [15]. The FRLP aims at locating a predetermined number of EV charging stations in a way to maximize charging demand satisfaction. Charging demand is represented by a set of round trips that drivers would like to carry out between pairs of nodes of the road network. A trip is considered 'covered' if drivers can travel from the origin of the trip to its destination and back without running out of charge.

Two main parameters have to be taken into account in order to evaluate the coverage of a trip: the battery energy status  $E$  when the EV leaves a charging station and its power consumption. A trip will be covered if, for each pair of consecutive stations visited when traveling along the corresponding path, the energy available in the battery when leaving the first station is high enough to provide the power needed by the vehicle to reach the second station. Two modeling assumptions are used in [15] and most of the subsequent works on the FRLP: (1) Each time an EV stops at a charging station, it is fully recharged, i.e.  $E$  is deterministically known and equal to the battery maximum capacity. (2) The unit power consumption per kilometer is deterministically known and has the same value on the entire road network: the power consumption is thus directly proportional to the distance traveled and does not depend on the crossing direction on any road segment. Under these two assumptions, the driving range, i.e. the distance that a fully charged vehicle can travel before its battery runs empty, which is computed as the ratio of the battery capacity over the unit power consumption, can be considered as deterministically known and constant.

However, in practice, the energy available in the battery when an EV leaves a charging station is influenced by many factors such as the age and technology of the battery and the time spent at the charging station. These factors may strongly vary from one vehicle to the other, making it difficult for the charging infrastructure planner to accurately forecast the expected energy status of an EV leaving a station. Moreover, the unit power consumption on a road segment is also subject to many uncertainties as it depends among others on the terrain configuration (road topology, road grade...) and on the driving conditions (e.g. road link travel speed, wind speed, status of on-board electric devices, driver's driving style...). As shown e.g. by the numerical tests provided by de Vries and Duijzer [6], not taking into account these uncertainties in the flow refueling location problem leads to under or overestimating the actual charging demand coverage so that the infrastructure provided by a deterministic model might be suboptimal. Therefore, explicitly considering the uncertainty of the driving range when modeling the problem is necessary to obtain better station location decisions.

The FRLP with a stochastic driving range was recently studied by de Vries and Duijzer [6], Kchaou Boujelben and Gicquel [13] and Lee and Han [17]. These works introduced the notion of coverage probability of a trip, which is defined as the probability that the drivers can carry out the trip without running out of fuel, i.e. as the joint probability that the power consumption between each pair of successively visited stations is smaller than the energy available in the battery when leaving a station. However, in order to be able to handle this joint probability in the problem formulation, these authors make use of rather strong assumptions on the corresponding random variables.

In the present work, we study the FRLP under uncertain battery energy status after leaving a charging station and uncertain power consumption. More precisely, we consider the chance constrained flow refueling location model proposed by de Vries and Duijzer [6]. In this model, a trip is considered covered if its coverage probability remains above a predefined minimum value and the objective is to maximize the covered flow. Our contributions can be summarized as follows:

- In contrast to most previously published works, we allow the power consumption on a road segment to depend on the crossing direction. This implies that checking whether a trip  $q$  is covered by the installed charging stations requires checking the coverage of all segments traveled by the EV both in the outward and return directions. As a consequence, the number of constraints to be introduced in the mixed-integer linear programming (MILP) formulation is doubled.
- We seek to relax the assumptions on the stochastic battery energy status when leaving a station and on the stochastic power consumption previously used in the literature to handle this problem. We namely consider a rather general case involving a random power consumption on each road segment, with statistical dependencies between the corresponding random variables, and a stochastic battery energy status after recharging. We only assume that the battery energy status after recharging is the same over the entire road network and is statistically independent of the power consumption on each road segment. This leads to the formulation of a mathematical program involving a set of joint chance constraints which are difficult to handle as such, especially if the random variables follow continuous probability distributions.
- As suggested in [6], we first implement a solution approach based on the approximation of each joint chance constraint using Bonferroni's inequality. This approach results in the formulation of a mixed-integer linear program of a size similar to the one of the deterministic problem. Moreover, the obtained solutions are guaranteed feasible solutions of the original joint chance-constrained program. However, as the approximation of the joint probability may largely underestimate the actual coverage probability of some trips, it may lead to very conservative solutions displaying a demand satisfaction much lower than what could have been obtained if the original stochastic problem had been solved to optimality.

- We thus propose a second solution approach based on a partial sample approximation (PSA) of the random variables. This technique is a recent extension proposed by Cheng, Gicquel and Lisser [4] of the sample approximation (SA) approach. Similarly to the SA approach introduced by Lüdtké and Ahmed [18], it relies on a Monte Carlo sampling of the continuous distribution of the random variables but this sampling is carried out on all random variables except one (the stochastic battery energy status after recharging in our case). The main advantage of the PSA approach is that, unlike the SA approach, it does not require the introduction of additional binary variables in the formulation. It thus leads to the formulation of a mixed-integer linear program (MILP) which has more continuous variables and constraints than the deterministic problem but the same number of binary variables as the deterministic problem.
- We provide numerical results obtained on medium-size randomly generated and real-life instances. These results show that the PSA approach is more computationally intensive than a solution approach based on Bonferroni’s approximation. However, it is capable of providing feasible solutions of the chance-constrained problem corresponding to an actual demand coverage significantly better than the one obtained with the Bonferroni’s approximation. As determining the location of EV charging stations is a strategic long-term problem, increasing the computation time by a reasonable amount to obtain better quality solutions seems a relevant trade-off.

The paper is organized as follows. An overview of the related literature is proposed in Section 2. Section 3 introduces the deterministic flow refueling location model (FRLM) with asymmetric power consumption. In Section 4, the chance constrained flow refueling location model under uncertain battery energy status and power consumption CCFRLM-EC is presented. Then, two solution approaches, the Bonferroni’s approximation and the partial sample approximation (PSA), are described in Section 5. Section 6 analyzes the results of the numerical experiments carried out to compare the performance of the proposed solution approaches on randomly generated instances as well as on real-life networks. Finally, Section 7 presents some conclusions and research perspectives.

## 2 Related literature

Facility location models used in the design of EV charging infrastructure can be classified into two main categories. The first category assumes that the demand is located at the nodes of a network while the second category represents the demand as a set of origin-destination trips. An example of a node-based model can be found in [8] where demand arises at the centroids of residential or workplace areas. Whether a fixed-charge location [20], a maximal-covering [8] or a set-covering model [29] is used, the problem modeling in that case involves the distance between each demand node and each charging station, either in the objective function or in the constraints of the resulting MILP formulation. Node-based models implicitly assume that EV drivers will specially drive from their work or house location to the charging station and back to recharge their battery. However, these models do not allow to represent the situation where drivers carry out long-distance trips and need to recharge their battery while traveling between a departure point and a destination in the road network. Flow-based models in which the recharging demand is modeled as a set of origin-destination trips taken by EV drivers should be more appropriate in this case. The flow capturing location model (FCLM) introduced by Hodgson [9] is among the first works to consider a flow-based model. Hodgson [9] assumes that an origin-destination flow is ‘captured’ if at least one facility is located anywhere on the shortest path between the origin and the destination of the corresponding flow. He formulates the problem using a maximal-covering model in which the objective is to maximize the number of flows captured while opening a limited number of facilities. The main issue in applying the FCLM for locating EV charging stations is that it does not take into account the limited range of an EV, i.e. the maximum distance that the EV can travel before its battery runs out of charge.

In order to fill this gap, Kuby and Lim [15] introduce the flow refueling location model (FRLM). In this model, a flow on a given path is considered as refueled if and only if there is an adequate number of stations on the path, allowing an EV driver to refuel his journey from the origin to the destination. Several works have recently studied the flow refueling location problem. For instance, Capar et al. [3] and MirHassani and Ebrazi [21] propose new mixed-integer linear programming formulations of the problem in order to solve it more efficiently with a mathematical programming solver. Upchurch and Kuby [28] study a capacitated version of the problem in which the assumption of unlimited charging capacity is relaxed. Another extension of the problem in which drivers are willing to select paths slightly deviating from the shortest path between their origin and their destination to charge their battery is investigated e.g. in [14] and [31]. However, most of these works assume a deterministic setting whereas in practice, the main parameters of the problem are subject to significant uncertainties. In what follows,

we thus focus on stochastic variants of the EV charging station location problem. We review the related works and provide a classification in Table 1.

Schiffer and Walther [27] and Davidov and Pantos [5] study node-based stochastic EV charging station location problems. In [27], a robust location routing problem under customer demand, spatial distribution and service time windows uncertainties is studied. To solve large instances of the problem, the authors combine an adaptive large neighborhood search algorithm with dynamic programming. In [5], a set-covering type model is developed for locating charging stations under driving range uncertainty. The range uncertainty is represented through a set of discrete scenarios and the deterministic demand for EV refueling is represented using a set of discrete points characterizing the dynamic location of EVs over time. The objective function minimizes the station placement cost while assigning a weight to each station, representing the flow of EVs, whose trajectories pass by the station.

The stochastic flow-based models presented in Table 1 involve a first set of studies (see [10], [12], [19], [30] and [32]) that focuses on taking into account in the problem modeling the uncertainties on the recharging demand. The paper of Mak et al. [19] has been among the first works to introduce uncertainties in this type of problem. They propose a distributionally robust optimization framework to design a battery swapping infrastructure under stochastic charging demand. In [12], Jung *et al.* investigate the problem of locating stations to recharge electric shared-ride taxis under stochastic demand and queuing delays. They propose a simulation-optimization approach and test it on a network in Seoul. Wu and Sioshansi [30] study a stochastic flow capturing location problem with uncertain vehicle flows and model it using a two-stage stochastic programming approach. Hosseini and MirHassani [10] extend the path-segment formulation they introduced in [21] to account for uncertain traffic flows. They propose a two-stage stochastic programming model in which portable recharging stations can be opened in the second decision stage in order to improve the expected flow coverage. Finally, a capacitated stochastic set-covering location problem with driver route deviation tolerance is studied by Yildiz *et al.* [32] and a novel branch and cut approach is developed to solve large instances of the problem.

Stochastic driver route choice behaviour is considered in [11] and [24], where the route between an origin node  $O$  and a destination node  $D$  is selected based on criteria such as the traffic congestion and the availability of charging facilities instead of being systematically assigned to the shortest path between  $O$  and  $D$ . In [24], a single level non linear model under stochastic user equilibrium (SUE) constraints is proposed while in [11], the authors develop a bi-level programming approach under uncertain driver route choice behaviour. The upper level is a maximum flow covering problem while the lower level deals with a stochastic traffic assignment problem under path distance constraints.

Ref.	Facility location		Stochastic parameters			Stochastic modeling		Optimization under uncertainty
	Demand	Model type	Demand	Range	Route choice	Other		Solution approach
[5]	N	Set covering		X			2SSP	MILP solver
[6]	F	FRLM		X			2SSP/CCP	MILP solver
[10]	F	FRLM	X				2SSP	Greedy heuristic
[11]	F	BLFL			X		SSSP	Equilibrium-based heuristic
[12]	F	BLFL	X			X	Queuing models	Optimization-simulation based heuristic
[13]	F	FRLM		X			2SSP/CCP	MILP solver and Tabu Search heuristic
[17]	F	FRLM		X			CCP	Benders and Price
[19]	F	FRLM					DRO	SOCP solver
[24]	F	FRLM	X		X		SSSP	Linearization approach and MILP solver
[27]	N	LRM	X			X	RO	Adaptive neighbourhood search
[30]	F	FCLM	X				2SSP	L-shaped algorithm
[32]	F	FRLM	X				2SSP	Branch and Cut
<b>This work</b>	<b>F</b>	<b>FRLM</b>					<b>CCP</b>	<b>PSA/Bonferromi with MILP solver</b>

Demand: N=node-based demand; F=flow-based demand

Model type: BLFLM = bi-level facility location model; FCLM= flow capturing location model ;

FRLM = flow refueling location model; LRM=location-routing model

Stochastic modeling: DRO=distributionally robust optimization; RO=robust optimization; CCP=chance-constrained programming ;

SSSP = single-stage stochastic programming; 2SSP = 2-stage stochastic programming

Solution approach: MILP= mixed-integer linear programming; SOCP= second-order cone programming

Table 1: Classification of the literature on stochastic problems

In most of the deterministic and stochastic demand flow-based models mentioned above, the coverage decision is binary, i.e. a trip is either covered or not, depending on the station location decisions. Yet, when the battery energy status after recharging and the power consumption are stochastic, it might not be possible anymore to ensure that a trip will be covered with a 100% probability, even if a large number of stations is opened on the corresponding path. Therefore, as explained in Section 1, the coverage probability of a trip, i.e. the joint probability that the power consumption on each portion of the trip comprised between two opened charging stations is smaller than the energy available in the battery after recharging, should be introduced in the model. To the best of our knowledge, only three recent works, [6], [13] and [17] have studied this problem. However, they all relied on rather strong assumptions on the random variables in order to be able to compute the coverage probability. Namely, de Vries and Duijzer [6] and Kchaou Boujelben and Gicquel [13] consider a single random variable representing the stochastic driving range over the entire road network. This accounts for the uncertainties related to the age and technology of the battery and to the general weather conditions. However, it ignores the uncertainties specific to each road segment, in particular those resulting from the local traffic and weather conditions. Lee and Han [17] introduce random variables modeling the stochastic driving range on each road link and assume that these random variables are all independent of one another. This assumption allows them to compute the coverage probability of a trip as the joint probability of independent events but the potential dependencies between the power consumption on various road links of the network cannot be taken into account. de Vries and Duijzer [6] propose to extend their model to consider correlated random variables modeling the stochastic driving ranges on each road link. They suggested a solution approach based on Bonferroni’s inequality to approximately compute the coverage probability of a trip in this case but did not provide numerical results to validate it.

In what follows, we study the FRLP with a stochastic battery energy status after recharging and power consumption under rather general assumptions on the related random variables. We investigate two approximate solution approaches for this joint chance-constraint problem, an approach based on the Bonferroni’s inequality and an approach based on a partial sample approximation (PSA), and provide a numerical assessment of their performance on a large set of medium-size randomly generated and real-life instances.

### 3 Deterministic Flow Refueling Location Problem with asymmetric power consumption

#### 3.1 Problem description

The road network is modeled by an oriented graph  $\mathcal{G}(\mathcal{N}, \mathcal{A})$ , where  $\mathcal{N}$  is the set of nodes and  $\mathcal{A}$  is the set of arcs in the network. The recharging demand is represented as a set of predetermined trips  $q \in \mathcal{Q}$ , where each trip  $q$  consists in traveling from an origin  $O^q \in \mathcal{N}$  to a destination  $D^q \in \mathcal{N}$  and back. We assume that all EV drivers follow the shortest path between the origin  $O^q$  and the destination  $D^q$  on trip  $q$  and do not deviate from it to recharge their battery. The flow on trip  $q$ , i.e. the number of vehicles traveling between  $O^q$  and  $D^q$  per unit of time, is denoted  $f^q$ . All problem notations are shown in Table 2.

We assume that all EVs, after charging at a station, have the same energy status or state of charge  $E$  (expressed e.g. in kWh). The problem is that the energy needed to carry out a trip  $q$  is likely to exceed  $E$  so that multiple stops at charging stations will be needed during the trip. Charging stations are assumed to be uncapacitated and accessible for traffic from both directions. In order for a trip  $q$  to be covered, i.e. for an EV driver to carry out his journey between  $O^q$  and  $D^q$  without running out of charge, a sufficient number of stations must be located on the shortest path between  $O^q$  and  $D^q$ . Contrary to most previously published works, we consider that the power consumption on a road segment may be asymmetric, i.e. may depend on the crossing direction. A trip  $q$  is thus considered covered if the power consumption between each pair of consecutive stations on  $q$  does not exceed  $E$  when traveling both in the outward ( $d = o$ ) and return ( $d = r$ ) directions. Otherwise, the trip is considered not covered. The objective is to determine the best locations for a predetermined number  $p$  of charging stations in a way to maximize the total flow of EVs covered.

In order to more easily model the fact that drivers cyclically carry out round trips between pairs of nodes in the network, de Vries and Duijzer [6] introduced the concept of cycle segment. A cycle segment  $[k, l]$  consists in the sequence of consecutive arcs to be crossed when traveling on the shortest path between node  $k$  and node  $l$ . However, when the segment starts at the origin of the trip (i.e.  $k = O^q$ ), there are two possibilities: either the driver travels in the outward direction from  $O^q$  to  $l$  or he carries out a round trip  $l \rightarrow O^q \rightarrow l$ . In order to differentiate between the two situations, we duplicate node  $O^q$  into node  $O'^q$  and add two arcs  $(O'^q, O^q)$  and  $(O^q, O'^q)$  of length 0 to  $\mathcal{A}$ : cycle segment  $[O^q, l]$  represents traveling in the outward direction from  $O^q$  to  $l$  and cycle segment  $[O'^q, l]$  represents the round trip to the origin. Similarly, we duplicate node  $D^q$  into node  $D'^q$  and add two arcs  $(D^q, D'^q)$

Parameter	Description
$\mathcal{N}$	Set of nodes of the road network
$\mathcal{A}$	Set of arcs $(i, j) \in \mathcal{N} \times \mathcal{N}$ of the road network
$\mathcal{Q}$	Set of origin-destination trips taken by EV drivers in the network
$E$	Expected energy status of an EV after charging at a station
$p$	Number of charging stations to be built
$f^q$	Flow of EVs on trip $q$
$O^q$	Origin of trip $q$
$O'^q$	Duplicated origin of trip $q$
$D^q$	Destination of trip $q$
$D'^q$	Duplicated destination of trip $q$
$\mathcal{N}^q$	Set of nodes belonging to the shortest path between $O'^q$ and $D'^q$
$\mathcal{N}^{qo}$	Set of nodes equal to $\mathcal{N}^q \setminus \{O'^q\}$
$\mathcal{N}^{qr}$	Set of nodes equal to $\mathcal{N}^q \setminus \{D'^q\}$
$\mathcal{N}_l^{qo}$	Set of nodes belonging to $\mathcal{N}^q$ visited before node $l$ when traveling in the outward direction on trip $q$
$\mathcal{N}_l^{qr}$	Set of nodes belonging to $\mathcal{N}^q$ visited before node $l$ when traveling in the return direction on trip $q$
$\mathcal{A}^q$	Set of arcs belonging to the shortest path between $O'^q$ and $D'^q$
$\mathcal{A}_{kl}^q$	Set of consecutive arcs visited when traveling on the shortest path between $k$ and $l$ during trip $q$
$t_{ab}$	Power consumption on arc $(a, b) \in \mathcal{A}$
$t_{kl}^q$	Power consumption on cycle segment $[k, l]$ on trip $q$
$\alpha$	Maximum acceptable level for the risk of running out of fuel during a trip

Table 2: Problem parameters

and  $(D'^q, D^q)$  of length 0 to  $\mathcal{A}$ . The power consumption  $t_{kl}^q$  on a cycle segment  $[k, l]$  of a trip  $q$  is defined in Table 3 for segments traveled in the outward direction. For segments traveled in the return direction, power consumption can be calculated in a similar way. Note that  $\mathcal{N}^q$  (resp.  $\mathcal{A}^q$ ) is the set of nodes (resp. arcs) on the shortest path between nodes  $O'^q$  and  $D'^q$ ,  $\mathcal{A}_{kl}^q$  is the set of consecutive arcs visited when traveling on the shortest path from  $k$  to  $l$  on trip  $q$  and  $t_{mn}$  is the power consumption on arc  $(m, n) \in \mathcal{A}$ . Since the duplicated nodes  $O'^q$  and  $D'^q$  are only used to model round trips on  $q$  and cannot be used to locate charging stations, cycle segment  $[O'^q, D'^q]$  is not allowed and its length is set to  $+\infty$  as shown in Table 3. This is in line with the assumption made by de Vries and Duijzer [6].

	$l = D'^q$	$l \neq D'^q$
$k = O'^q$	$+\infty$	$\sum_{(m,n) \in \mathcal{A}_{kl}^q} t_{mn} + \sum_{(m,n) \in \mathcal{A}_{lk}^q} t_{mn}$
$k \neq O'^q$	$\sum_{(m,n) \in \mathcal{A}_{kl}^q} t_{mn} + \sum_{(m,n) \in \mathcal{A}_{lk}^q} t_{mn}$	$\sum_{(m,n) \in \mathcal{A}_{kl}^q} t_{mn}$

Table 3: Power consumption on segment  $[k, l]$  when traveling in the outward direction on trip  $q$

### 3.2 Mathematical formulation

An efficient formulation of the flow refueling location problem (FRLP) with symmetric power consumption has been recently proposed in [13]. This formulation relies on the idea that, in order to cover a trip  $q$ , we should assign to each node  $l \in \mathcal{N}^q$  a single station  $k \in \mathcal{N}_l^{qo}$  located within the driving range of the vehicle to enable the EV driver to recharge his battery at station  $k$  and travel at least up to node  $l$ . In case trip  $q$  is not covered, there is no need to assign a charging station to each node  $l \in \mathcal{N}^q$ . Note that this formulation focuses only on making the trip feasible in the outward direction as, thanks to the assumption of symmetric power consumption, this is sufficient to ensure that the round trip is feasible.

We propose in this section to extend this formulation to the case of asymmetric power consumption. Accordingly, for a trip  $q$  to be covered, we should assign to each node  $l \in \mathcal{N}^q$  a single station  $k \in \mathcal{N}_l^{qo}$  in charge of serving it when traveling in the outward direction of  $q$  and a single station  $k' \in \mathcal{N}_l^{qr}$  in charge of serving it when traveling in the return direction of  $q$ . Both stations should be close enough to ensure that  $E$  is larger than  $t_{kl}^q$  (resp.  $t_{k'l}^q$ ) to enable the EV driver to refuel his vehicle at node  $k$  (resp.  $k'$ ) and reach node  $l$  in the outward (resp. return) direction.

We first introduce two sets of binary variables defined by:

- $x_k = 1$  if a station is built at node  $k$ , 0 otherwise.
- $y^q = 1$  if the trip  $q$  is covered, 0 otherwise.

We also introduce binary assignment variables  $w_{kl}^q$  defined as follows:



$$w_{kl}^q = \begin{cases} 1 & \text{- if we assign to node } l \in \mathcal{N}^{qd} \text{ a station located at node } k \in \mathcal{N}_l^{qd} \setminus \{O^q, D^q\} \text{ to refuel the driver} \\ & \text{journey in direction } d \text{ at least up to node } l \\ & \text{- if } k = O^q \text{ (respectively } k = D^q) \text{ and the subtrip } l \rightarrow O^q \rightarrow l \text{ (respectively } l \rightarrow D^q \rightarrow l) \\ & \text{is refueled by a station located at node } l \in \mathcal{N}^{qo} \text{ (respectively } \mathcal{N}^{qr}) \text{ or before it when traveling in} \\ & \text{the return (respectively outward) direction} \\ 0 & \text{otherwise} \end{cases}$$

The FRLP with asymmetric power consumption can be formulated as follows:

$$\max \quad \sum_{q \in \mathcal{Q}} f^q y^q \quad (1)$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{N}_l^{qd}} w_{kl}^q = y^q \quad \forall q \in \mathcal{Q}, d \in \{o, r\}, l \in \mathcal{N}^{qd} \quad (2)$$

$$w_{kl}^q \leq x_k \quad \forall q \in \mathcal{Q}, d \in \{o, r\}, l \in \mathcal{N}^{qd}, k \in \mathcal{N}_l^{qd} \setminus \{D^q, O^q\} \quad (3)$$

$$\sum_{k \in \mathcal{N}_l^{qd}} t_{kl}^q w_{kl}^q \leq E \quad \forall q \in \mathcal{Q}, d \in \{o, r\}, l \in \mathcal{N}^{qd} \quad (4)$$

$$\sum_{j \in \mathcal{N}} x_j = p \quad (5)$$

$$x_k, y^q, w_{kl}^q \in \{0, 1\} \quad (6)$$

The objective function maximizes the flow of EVs covered computed as the sum of covered trips weighted by the flow of EVs on each trip. Constraints (2) are assignment constraints ensuring that, if trip  $q$  is covered, each node  $l \in \mathcal{N}^{qo}$  (respectively  $l \in \mathcal{N}^{qr}$ ) has to be assigned to a single node  $k$  situated before  $l$  when traveling in the outward direction (respectively return direction) of the trip. Note how these constraints set to 0 all assignment variables  $w_{kl}^q$  corresponding to trip  $q$  in case it is not covered. Constraints (3) state that if no station is opened at node  $k$ , drivers cannot refuel their battery at node  $k$  and all variables  $w_{kl}^q$  corresponding to node  $k$  are set to 0. Constraints (4) guarantee that for a trip  $q$  to be covered, the power consumption between each pair of successively visited stations is smaller than the energy available in the battery after recharging. More precisely, they ensure that, for each node  $l$  on path  $q$ , the power consumption between the node  $k$  to which  $l$  is assigned and node  $l$  is smaller than  $E$ , both in the outward ( $d = o$ ) and return ( $d = r$ ) directions. Note that, in case trip  $q$  is not covered, all  $w_{kl}^q$  variables corresponding to trip  $q$  are set to 0 and constraints (4) become inactive. Constraint (5) is the budget constraint: it limits the number of charging stations that can be opened to a predefined number  $p$ . Finally, constraints (6) define the problem variables.

## 4 Stochastic flow refueling location problem under uncertain battery energy status and power consumption

As explained in Sections 1 and 2, when the energy status of the EV battery after recharging and/or the power consumption on each road segment are subject to uncertainties, trip coverage becomes a matter of chance rather than of binary observation. We thus introduce the notion of coverage probability of a trip defined as the joint probability that the power consumption between any pair of consecutive stations visited by the EV driver in the outward and return directions of the trip be smaller than the amount of energy available in the battery after recharging. Furthermore, as suggested by de Vries and Duijzer in [6], an EV driver might not be willing to take a trip if the probability of running out of charge during the trip is above a maximum acceptable risk level  $\alpha$ . De Vries and Duijzer thus propose the chance constrained flow refueling location model in which a trip is considered covered if its coverage probability is higher than  $1 - \alpha$ , and not covered otherwise. In what follows, we study the chance constrained flow refueling location problem under uncertain battery energy status and power consumption denoted CCFRLM-EC.

Let  $\tilde{E}$  be the random variable representing the stochastic energy status of the battery after visiting a station and  $\tilde{t}_{kl}^q$  be the random power consumption on segment  $[k, l]$  of path  $q$ .

The coverage variables  $y^q$  are now defined by  $y^q = 1$  if the coverage probability of trip  $q$  is above  $1 - \alpha$  and 0 otherwise.

This leads to the following formulation denoted CCFRLM-EC:

$$\max \quad \sum_{q \in \mathcal{Q}} f^q y^q \quad (7)$$

$$\text{s.t.} \quad P \left( \sum_{k \in \mathcal{N}_l^{qd}} \widetilde{t}_{kl}^q w_{kl}^q \leq \widetilde{E} \quad \forall d \in \{o, r\}, l \in \mathcal{N}^{qd} \right) \geq 1 - \alpha \quad \forall q \in \mathcal{Q} \quad (8)$$

(2) – (3), (5) – (6)

We first note that in case path  $q$  is not covered, all variables  $w_{kl}^q$  are equal to 0 (thanks to constraints (2)) so that the corresponding joint chance constraint (8) is satisfied. Moreover, in case trip  $q$  is covered, for each node  $l \in \mathcal{N}^{qo}$ , there exists a single node  $k_l \in \mathcal{N}_l^{qo}$  such that  $w_{k_l l}^q = 1$ . Similarly, for each node  $l \in \mathcal{N}^{qr}$ , there exists a single node  $k'_l \in \mathcal{N}_l^{qr}$  such that  $w_{k'_l l}^q = 1$ . Constraints (8) thus imposes that the probability of not running out of fuel when traveling in the outward direction from the refueling station at node  $k_l$  to node  $l$  and in the return direction from the refueling station at node  $k'_l$  to node  $l$ , for each node  $l$  belonging to the shortest path, be higher than  $1 - \alpha$ .

As explained e.g. by Luedtke and Ahmed [18], most mathematical programs involving a joint chance constraint are still largely computationally intractable. There are two main reasons for this. First, given a solution of the mathematical program, checking its feasibility, i.e. checking whether it complies with the joint chance constraint, requires to evaluate a multidimensional integral to compute the value of the joint probability, which can be very time-consuming. Second, in general, the feasible set defined by a joint chance-constraint is not convex. Yet, we need both conditions, i.e. efficient computation of the joint probability and convexity of the solution space, to efficiently process chance-constraints: see e.g. Nemirovski and Shapiro [22]. In view of these difficulties, a variety of tractable approximations have been proposed for joint chance-constrained problems. A first set of approaches is based on conservative convex approximations of the feasible space: see e.g. Bonferroni [1], Rockafellar and Uryasev [25] and Nemirovski and Shapiro [23]. A second set uses a discretization of the probability distribution through sampling to approximately compute the joint probability: see e.g. Calafiore and Campi [2], Luedtke and Ahmed [18], Cheng, Gicquel and Lisser [4].

In what follows, we investigate two of these approaches: a conservative approximation based on Bonferroni's inequality [1] and the partial sample approximation proposed in [4].

## 5 Solution approaches

### 5.1 Bonferroni approximation

As suggested by de Vries and Duijzer in [6], a first approach to deal with the joint chance constraints (8) relies on Bonferroni's inequality [1].

Let  $R_l^{qd}$  be the random event that the EV successfully reach node  $l$  when traveling on direction  $d \in \{o, r\}$  along trip  $q$  after recharging at a station located at a node  $k \in \mathcal{N}_l^{qd}$ , i.e. the event  $\sum_{k \in \mathcal{N}_l^{qd}} \widetilde{t}_{kl}^q w_{kl}^q \leq \widetilde{E}$ .

With this notation, constraints (8) can be rewritten as:

$$P \left( \bigcap_{d \in \{o, r\}, l \in \mathcal{N}^{qd}} R_l^{qd} \right) \geq 1 - \alpha \quad \forall q \in \mathcal{Q}.$$

Thanks to Bonferroni's inequality, we have:

$$P \left( \bigcap_{d \in \{o, r\}, l \in \mathcal{N}^{qd}} R_l^{qd} \right) \geq 1 - \sum_{d \in \{o, r\}} \sum_{l \in \mathcal{N}^{qd}} P \left( \overline{R_l^{qd}} \right) \quad \forall q \in \mathcal{Q}.$$

By replacing the joint probability by its lower bound, we obtain the following conservative approximation of constraints (8):

$$\sum_{d \in \{o, r\}} \sum_{l \in \mathcal{N}^{qd}} P \left( \sum_{k \in \mathcal{N}_l^{qd}} \widetilde{t}_{kl}^q w_{kl}^q \geq \widetilde{E} \right) \leq \alpha \quad \forall q \in \mathcal{Q} \quad (9)$$

We note that constraints (9) involve a sum of terms, each one corresponding to the probability of the event that the EV runs out of fuel before reaching node  $l$  when traveling in direction  $d$ . This probability corresponds to the

probability of a single random event and is thus much easier to compute than the probability of the intersection of a set of random events. In particular, we can exploit the specific structure of the random constraints  $\sum_k \widetilde{t}_{kl}^q w_{kl}^q \geq \widetilde{E}$  to further simplify the computation. Namely, in any optimal solution,  $\forall l \in \mathcal{N}^{qo}$ , there is at most one node  $k \in \mathcal{N}_l^{qo}$  such that  $w_{kl}^q = 1$ . Similarly,  $\forall l \in \mathcal{N}^{qr}$ , there is at most one node  $k \in \mathcal{N}_l^{qr}$  such that  $w_{kl}^q = 1$  (see [13]). This enables us to reformulate constraints (9) as:

$$\sum_{d \in \{o,r\}} \sum_{l \in \mathcal{N}^{qd}} \sum_{k \in \mathcal{N}_l^{qd}} w_{kl}^q P\left(\widetilde{t}_{kl}^q \geq \widetilde{E}\right) \leq \alpha \quad (10)$$

Let  $g_{kl}^q = P(\widetilde{t}_{kl}^q \geq \widetilde{E})$  be the probability that a vehicle recharging at a station located at node  $k$  runs out of fuel before reaching node  $l$ , i.e. the probability that the power consumption on cycle segment  $[k, l]$  be larger than the energy available in the battery after recharging at node  $k$ . The value of  $g_{kl}^q$  can be computed in a pre-optimization step as shown in A.

This leads to the following MILP formulation denoted CCFRLM-B:

$$\max \sum_{q \in \mathcal{Q}} f^q y^q \quad (11)$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{d \in \{o,r\}} \sum_{l \in \mathcal{N}^{qd}} \sum_{k \in \mathcal{N}_l^{qd}} g_{kl}^q w_{kl}^q \leq \alpha & \forall q \in \mathcal{Q} & (12) \\ & (2) - (3), (5) - (6) \end{aligned}$$

As mentioned in Section 1, this reformulation has two main advantages. First, it involves the same number of variables and constraints as the original stochastic problem. Second, as the joint probability has been replaced by a lower bound in the approximation of each chance constraint, a feasible solution of CCFRLM-B is guaranteed to comply with the original joint chance constraints (8) and thus to be a feasible solution of the original stochastic problem CCFRLM-EC. However, the lower bound provided by Bonferroni's inequality might be significantly smaller than the actual value of the joint probability, which means that the probability of successfully carrying out a trip might be underestimated in some cases. Constraints (12) might thus significantly reduce the feasible space of the problem so that the optimal solution of CCFRLM-B may be rather conservative and its actual performance in terms of demand coverage might be significantly lower than what an optimal solution of problem CCFRLM-EC would have provided.

## 5.2 Partial sample approximation

We thus investigate a second solution approach based on the partial sample approximation (PSA) proposed by Cheng, Gicquel and Lissner [4]. The main idea of this approach consists in replacing the original continuous probability distribution of all but one random parameters by an empirical discrete finite probability distribution obtained by Monte Carlo sampling. In the present case, we use a discrete approximation for the random power consumption on each road segment but keep a continuous probability distribution for the energy status of the battery after recharging. This leads to the formulation of a new stochastic program comprising a larger number of joint chance constraints than the original stochastic problem. However, each of these new joint chance constraints involves a single random variable and will thus be much easier to handle than the original joint chance constraints. Moreover, constraints (4) display a special structure which can also be exploited. This will enable us to reformulate problem CCFRLM-EC as a mixed-integer linear program involving the same number of binary variables as problem CCFRLM-EC.

In order to be able to use the partial sample approximation approach, we introduce the following assumption on the random parameters.

*Assumption:* For any cycle segment  $[k, l]$  belonging to trip  $q \in \mathcal{Q}$ , the power consumption on  $[k, l]$ ,  $\widetilde{t}_{kl}^q$ , is statistically independent of the battery energy status after recharging  $\widetilde{E}$ .

Note that this assumption may seem reasonable since most of the sources of uncertainty of the battery energy status after recharging (e.g. age of the battery, technology of the battery, time spent in the charging station) differ from the sources of the power consumption randomness (e.g. heavy traffic).

Let  $t_{kl}^{q1}, \dots, t_{kl}^{qs}, \dots, t_{kl}^{qS}$  be a Monte Carlo sampling of the random parameter  $\widetilde{t}_{kl}^q$ . The  $S$  sampled scenarios are independent and identically distributed observations of vector  $\widetilde{t}_{kl}^q$ . The probability of each scenario  $s$  is thus set to  $\Pi_s = \frac{1}{S}$ .

Now, let  $R^q$  be the random event that trip  $q$  is covered, i.e.  $R^q = \left[ \sum_{k \in \mathcal{N}_l^{qd}} \widetilde{t}_{kl}^q w_{kl}^q \leq \widetilde{E} \quad \forall d \in \{o, r\}, l \in \mathcal{N}^{qd} \right]$ . We denote  $1_{R^q}$  the indicator function of this event:  $1_{R^q}$  is equal to one if the event has occurred and zero otherwise.

With this partial sampling of the random parameters, given a node-station assignment  $w^q$ , the coverage probability of trip  $q$ ,  $P(R^q)$ , can be approximately computed as follows:

$$P(R^q) = P \left( \sum_{k \in \mathcal{N}_l^{qd}} \widetilde{t}_{kl}^q w_{kl}^q \leq \widetilde{E} \quad \forall d \in \{o, r\}, l \in \mathcal{N}_d^q \right) \quad (13)$$

$$= \mathbb{E}_{(\widetilde{t}^q, \widetilde{E})} [1_{R^q}] \quad (14)$$

$$= \mathbb{E}_{\widetilde{t}^q} \mathbb{E}_{\widetilde{E}} [1_{R^q}] \quad (15)$$

$$= \mathbb{E}_{\widetilde{t}^q} \left[ P \left( \sum_{k \in \mathcal{N}_l^{qd}} \widetilde{t}_{kl}^q w_{kl}^q \leq \widetilde{E} \quad \forall d \in \{o, r\}, l \in \mathcal{N}_d^q \right) \right] \quad (16)$$

$$\simeq \frac{1}{S} \sum_{s=1}^S P \left( \sum_{k \in \mathcal{N}_l^{qd}} t_{kl}^{qs} w_{kl}^q \leq \widetilde{E} \quad \forall d \in \{o, r\}, l \in \mathcal{N}_d^q \right) \quad (17)$$

$$\simeq \frac{\sum_{s=1}^S v^{qs}}{S} \quad (18)$$

Equalities (14) and (16) make use of the fact that the probability of the random event  $R^q$  is equal to the expected value of the indicator function  $1_{R^q}$ . The assumption that  $\widetilde{E}$  is statistically independent of  $\widetilde{t}^q$  is exploited in equality (15) to decompose the computation of the expected value in two parts. A sample average approximation based on the sample of  $S$  scenarios of the stochastic power consumption  $\widetilde{t}^q$  is then used in (17) to approximately compute the expected value involved in equality (16). By defining  $v^{qs} = P \left( \sum_{k \in \mathcal{N}_l^{qd}} t_{kl}^{qs} w_{kl}^q \leq \widetilde{E} \quad \forall d \in \{o, r\}, l \in \mathcal{N}_d^q \right)$ , we obtain (18).

Intuitively, the partial sample approximation approach may be understood as follows. For each trip  $q$  and each scenario  $s$ , we compute  $v^{qs}$ , the probability that trip  $q$  is covered in scenario  $s$ , i.e. the probability that the power consumption on each segment  $[k, l]$  of trip  $q$  under scenario  $s$  stays below the random energy battery status after recharging  $\widetilde{E}$ . The value of the joint probability  $P(R^q)$  is then approximately estimated by the expected value of  $v^{qs}$  over all scenarios, i.e. by  $\frac{\sum_{s=1}^S v^{qs}}{S}$ .

We note that the definition of  $v^{qs}$  still involves the joint probability of a set of random events but as this joint probability now involves a single random variable  $\widetilde{E}$ , the value of  $v^{qs}$  can be computed quite easily by noting that:

$$v^{qs} = P \left( \max_{d \in \{o, r\}, l \in \mathcal{N}^{qd}} \sum_{k \in \mathcal{N}_l^{qd}} t_{kl}^{qs} w_{kl}^q \leq \widetilde{E} \right) \quad (19)$$

In other words,  $v^{qs}$  represents the probability that, under scenario  $s$ , the highest energy consumption of the EV on a cycle segment of trip  $q$  be less than the energy available after recharging.

Let  $G_{\widetilde{E}}$  be the cumulative density function of the random variable  $\widetilde{E}$ .

We have:

$$v^{qs} = 1 - G_{\tilde{E}} \left( \max_{d \in \{o,r\}, l \in \mathcal{N}^{qd}} \sum_{k \in \mathcal{N}_l^{qd}} t_{kl}^{qs} w_{kl}^q \right) \quad (20)$$

$$= 1 - \max_{d \in \{o,r\}, l \in \mathcal{N}^{qd}} G_{\tilde{E}} \left( \sum_{k \in \mathcal{N}_l^{qd}} t_{kl}^{qs} w_{kl}^q \right) \quad (21)$$

$$= 1 - \max_{d \in \{o,r\}, l \in \mathcal{N}^{qd}} \sum_{k \in \mathcal{N}_l^{qd}} G_{\tilde{E}}(t_{kl}^{qs}) w_{kl}^q \quad (22)$$

$$= 1 - \max_{d \in \{o,r\}, l \in \mathcal{N}^{qd}} \sum_{k \in \mathcal{N}_l^{qd}} \gamma_{kl}^{qs} w_{kl}^q \quad (23)$$

Note that equality (21) holds thanks to the fact that  $G_{\tilde{E}}$  is a non-decreasing function. Moreover, in any feasible solution of the problem,  $\forall d \in \{o,r\}, \forall l \in \mathcal{N}^{qd}$ , there is at most one node  $k \in \mathcal{N}_l^{qd}$  such that  $w_{kl}^q = 1$ : see constraints (2). We thus have  $G_{\tilde{E}} \left( \sum_{k \in \mathcal{N}_l^{qd}} t_{kl}^{qs} w_{kl}^q \right) = \sum_{k \in \mathcal{N}_l^{qd}} G_{\tilde{E}}(t_{kl}^{qs}) w_{kl}^q$ , which gives equality (22). By denoting  $\gamma_{kl}^{qs} = G_{\tilde{E}}(t_{kl}^{qs})$  the probability that  $\tilde{E}$  is smaller than the power consumption on segment  $[k, l]$  in scenario  $s$ , we obtain equality (23).

This leads to the following MILP formulation denoted by CCFRLM-D:

$$\max \sum_{q \in \mathcal{Q}} f^q y^q \quad (24)$$

$$s.t. \ v^{qs} \leq 1 - \sum_{k \in \mathcal{N}_l^{qd}} \gamma_{kl}^{qs} w_{kl}^q \quad \forall s = 1 \dots S, q \in \mathcal{Q}, d \in \{o,r\}, l \in \mathcal{N}^{qd} \quad (25)$$

$$\frac{\sum_{s \in S} v^{qs}}{S} \geq 1 - \alpha \quad \forall q \in \mathcal{Q} \quad (26)$$

$$v^{qs} \in [0, 1] \quad \forall s \in 1 \dots S, q \in \mathcal{Q} \quad (27)$$

$$(2) - (3), (5) - (6)$$

Formulation CCFRLM-D is based on an approximate representation of the joint chance constraints (8). As a consequence, there is no definite guarantee that a feasible solution of CCFRLM-D will also be feasible for the original stochastic problem as it may violate the joint chance constraints (8). However, Cheng, Gicquel and Lisser [4] showed that the probability that problem CCFRLM-D provides an optimal solution to the original stochastic problem approximates to 1 when the number of scenarios  $S$  goes to infinity.

## 6 Numerical experiments

We present in this section the results of computational experiments carried out on medium size randomly generated instances as well as on instances based on real-life networks reported in the literature. The main objective of these experiments is to assess the effectiveness of the proposed partial sample approximation approach by comparing it with Bonferroni's conservative approximation: see Subsection 6.3 for the results obtained with the randomly generated instances and Subsection 6.4.3 for the results obtained on with real-life networks. We also aim at analyzing the impact on the solution quality of using simplifying modeling assumptions such as a symmetric power consumption and deterministic input data. This is done by focusing on the instances based on real-life networks. For these instances, we first assess the impact on the demand coverage of assuming a symmetric power consumption when this latter is in fact asymmetric (see Subsection 6.4.1). We then seek to estimate the value of the stochastic solution by comparing its performance in terms of demand coverage with the one of the deterministic solution (see Subsection 6.4.2).

## 6.1 Test instances

### 6.1.1 Randomly generated instances

**Road network** Our procedure to randomly generate road networks is similar to the one proposed by Capar et al. in [3]. The first step is to randomly generate the coordinates of  $|\mathcal{N}|$  nodes according to a uniform distribution within  $[1, 1000]^2$ . The expected power consumption  $\bar{t}_{ab}$  to travel from node  $a$  to node  $b$  is computed as the product of the Euclidean distance between  $a$  and  $b$ , the unit power consumption (set to 0.2kWh per km) and a coefficient randomly generated between 0.8 and 1.2 to account for a potential road gradient. Note that we generate a different coefficient to obtain the expected power consumption  $\bar{t}_{ba}$  to travel from  $b$  to  $a$ . We then add to the network all arcs belonging to the minimum spanning tree. However, contrary to [3], we have a directed graph. We thus build an undirected graph by assigning to each pair of nodes  $(a, b)$  a unique weight equal to  $\frac{\bar{t}_{ab} + \bar{t}_{ba}}{2}$  and run Kruskal’s algorithm to find the minimum spanning tree comprising  $|\mathcal{N}| - 1$  arcs. We then add to  $\mathcal{A}$  the  $|\mathcal{N}|$  shortest potential arcs of the undirected graph not yet added to  $\mathcal{A}$  and such that no node of the undirected graph has a degree above four. Finally, for each arc  $(a, b)$  added to  $\mathcal{A}$ , we also add arc  $(b, a)$  in order to obtain a directed graph comprising a total of  $2(2|\mathcal{N}| - 1)$  arcs.

**Trips** The set  $\mathcal{Q}$  of trips to be covered is built by randomly selecting  $M$  origin-destination nodes among the  $|\mathcal{N}|$  nodes of the network, which results in  $|\mathcal{Q}| = \frac{M(M-1)}{2}$  trips to be covered. For each trip  $q \in \mathcal{Q}$ , the shortest path  $(\mathcal{N}^q, \mathcal{A}^q)$  between  $O^q$  and  $D^q$  is determined using Dijkstra algorithm. We assume that drivers use the same path, i.e. visit the set of nodes  $\mathcal{N}^q$ , when traveling on the return direction from  $D^q$  to  $O^q$ . The population at each origin/destination node is randomly generated within  $[1, 10000000]$  according to a uniform distribution. The flow  $f^q$  of EVs on each trip  $q$  is determined using the gravity model (see [7]). However, since long distance routes constitute the main focus of our paper, we change the formula as suggested in [6] in order to reduce the impact of short distance routes on location decisions. The flow on a path  $q$  is thus calculated as:  $f^q = \frac{P_O^q P_D^q}{t_{O^q, D^q}} \mathbf{1}_{t_{O^q, D^q} \geq 100}$  where  $P_O^q$  and  $P_D^q$  are the populations at the origin and destination nodes, respectively, and  $t_{O^q, D^q}$  is the length of the shortest path between  $O^q$  and  $D^q$ .

**Random parameters and maximum acceptable risk level** The random energy status after recharging  $\tilde{E}$  is assumed to follow a normal distribution, with an expectation  $\bar{E}$  set to 50kWh and a standard deviation  $\sigma_E$  set to  $\varphi \bar{E}$ . The random vector  $\tilde{t}$  describing the power consumption on each arc  $\tilde{t}_{ab} \in \mathcal{A}$  is assumed to follow a multivariate normal distribution characterized by an expected consumption vector  $\bar{t}$  and a covariance matrix  $\sigma$ . The expected value of  $\tilde{t}_{ab}$  is set to  $\bar{t}_{ab}$  and its standard deviation to  $\sigma_{ab} = \varphi \bar{t}_{ab}$ . The non-diagonal terms of the covariance matrix  $\sigma$ , representing the dependencies between the power consumption on different arcs of the network, are obtained by randomly generating the correlation matrix using a simple procedure that ensures that the obtained matrix is positive semi-definite (see e.g. [26]). The random consumption on cycle segment  $[k, l]$  is the sum of the random consumption of each arc belonging to the segment.  $\tilde{t}_{kl}^q$  is thus equal to the sum of a set of dependent random variables, each one following a normal distribution, and is therefore also normally distributed. Its expected value  $\bar{t}_{kl}^q$  and standard deviation  $\sigma_{kl}$  are computed as shown in A.

Finally, the maximum acceptable risk level  $\alpha$ , is set to 5 %.

**Instance sets** We consider two problem dimensions:  $(|\mathcal{N}| = 80, M = 40)$  and  $(|\mathcal{N}| = 100, M = 50)$ , corresponding respectively to 780 and 1225 trips. For each problem dimension, we randomly generate 10 networks and sets of trips to be covered. We consider 2 possible values of the number of stations:  $p \in \{5, 25\}$  and two possible values of the relative standard deviation:  $\varphi \in \{10\%, 20\%\}$ . This leads to a set of 80 test cases. The 10 test cases corresponding to the same problem dimension are grouped in the same set, resulting in 8 sets. Each set is denoted by providing the corresponding values of  $|\mathcal{N}|$ ,  $M$ ,  $p$ , and  $\varphi$ . For example, set 100.50.5.20 corresponds to the 10 test cases involving a network of 100 nodes, 50 origin/destination nodes, 5 stations to be opened and a relative standard deviation of 20%. The size of the Monte Carlo sample used in the partial sample approximation approach is set to  $S = 500$ . The discrete power consumption scenarios are obtained thanks to the Matlab procedure *mvnrnd* which returns a predefined number of vectors randomly sampled from a multivariable normal distribution. For each of the 80 test cases, 5 samples are generated.

We thus have a total of 80 instances to be solved with formulation CCFRLM-B and 400 instances to be solved with formulation CCFRLM-D.

### 6.1.2 Real-life instances

In our numerical experiments, we also use two real life case studies reported in the literature: the first one is located in California (see [31]), the second one in Florida (see [16]).

**Road networks** The California network consists of 339 nodes and 1234 arcs while the Florida network consists of 302 nodes and 848 arcs. In both cases, the distance matrix is symmetric. In order to obtain the average power consumption between each pair of nodes  $(a, b)$ , we multiply the distance between  $a$  and  $b$  by an average unit power consumption of 0.2kWh per km. The power consumption on arc  $a \rightarrow b$ ,  $t_{ab}$ , and the one on arc  $b \rightarrow a$ ,  $t_{ba}$ , are obtained by multiplying the average power consumption by a coefficient randomly generated between 0.8 and 1.2 to account for a potential difference in the power consumption in the outward and return directions on a road arc.

**Trips** In the California instance, 50 urban centers with a population over 50000 persons are used as origin-destination nodes and a set of 1167 paths is considered. The Florida instance involves 74 origin-destination nodes corresponding to rural counties and urban areas, leading to a total of 2701 paths. In both cases, the flow on each path is computed using the gravity model.

**Random parameters and maximum acceptable risk level** The random energy status after recharging  $\tilde{E}$  is assumed to follow a normal distribution, with an expectation  $\bar{E}$  set to 20kWh and a standard deviation  $\sigma_E$  set to  $\varphi\bar{E}$ . The random vector  $\tilde{t}$  describing the power consumption on each arc  $\tilde{t}_{ab} \in \mathcal{A}$  is assumed to follow a multivariate normal distribution characterized by an expected consumption vector  $\bar{t}$  and a covariance matrix  $\sigma$ .  $\sigma$  is built using the same procedure as the one used for the randomly generated instances.

**Instance sets** For each studied network, we consider two values for the number of stations  $p$ : 5 and 25 and two values for the relative standard deviation  $\varphi \in \{10\%, 20\%\}$ , which leads to a total of 8 test cases. Each test case is referred to as  $Net.p.\varphi$  where  $Net \in \{CA, FL\}$  relates to the network,  $p$  to the number of stations to be opened and  $\varphi$  to the relative standard deviation.

The size of the Monte Carlo sample used in the partial sample approximation approach is set to  $S = 100$  scenarios. The sampling is done using the same Matlab procedure as the one used for the randomly generated instances. For each test case, 5 samples are generated, leading to a total of 8 instances to be solved with formulation CCFRLM-B and 40 instances to be solved with formulation CCFRLM-D.

## 6.2 Experimental setup

Each of the 88 test cases was solved with the Bonferoni's conservative approach and the PSA approach. The MILP models CCFRLM-B and CCFRLM-D were implemented using the Concert Technology and solved by IBM ILOG CPLEX 12.6.2 with the solver default settings. We set the CPU time limit to 10 hours. All tests were carried out on a PC with Intel Core i7-8550U (1.80 GHz 1.99 GHz) with 16GB of RAM, running under Windows 10.

As explained in Section 5, both solution approaches rely on an approximation of the joint chance constraints and there is no guarantee that they will provide an optimal solution of problem CCFRLM-EC. We thus carry out a post-optimization analysis based on simulation in order to evaluate the quality of the charging infrastructure provided by the optimal solutions of models CCFRLM-B and CCFRLM-D. This simulation is carried out for each instance as follows.

Given  $x^*$  the optimal station locations provided by the MILP model,

1. We randomly generate a large number  $S'$  of scenarios for the battery energy status after recharging and the power consumption on each arc.
2. For each trip  $q$  and each scenario  $s'$ , given the values of  $E^{s'}$  and  $t_{kl}^{qs'}$  and the stations opened on path  $\mathcal{N}^q$ , we compute the coverage probability of trip  $q$  in scenario  $s'$ :  $\pi^{qs'}$ .
3. The probability that trip  $q$  is covered is then computed as the expected coverage probability of  $q$  over all scenarios  $s' = 1..S'$ , i.e. as  $\pi^q = \frac{\sum_{s'=1}^{S'} \pi^{qs'}}$ .
4. If  $\pi^q$  is higher than  $1 - \alpha$ , the coverage variable  $\hat{y}^q$  is set to 1. Otherwise,  $\hat{y}^q$  is set to 0.
5. The actual EV flow coverage provided by  $x^*$  is then estimated as  $\hat{Z} = \sum_{q \in \mathcal{Q}} f^q \hat{y}^q$ .

In our numerical experiments,  $S'$  was set to 5000 for the randomly generated instances and to 1000 for the real-life instances.

For each instance, the relative difference between the actual EV flow coverage obtained using each of the two solution methods is calculated as follows:  $\Delta = \frac{\hat{Z}_D - \hat{Z}_B}{\hat{Z}_B}$  where  $\hat{Z}_D$  (resp.  $\hat{Z}_B$ ) is the actual flow coverage obtained by simulation for the CCFRLM-D (resp. CCFRLM-B) location solution. For the randomly generated instances, the average value of  $\Delta$  over all instances belonging to the same set,  $\bar{\Delta}$ , is reported in Table 4 and the detailed results obtained for each of the considered 80 test cases are provided in Tables 10 to 17 (see B). Tables 8 and 9 display the detailed results for each studied test case corresponding to real-life networks.

## 6.3 Results for the randomly generated instances

### 6.3.1 Overall analysis

Set	100.50.5.10	100.50.25.10	100.50.5.20	100.50.25.20	80.40.5.10	80.40.25.10	80.40.5.20	80.40.25.20	Avg.
CPU-B(s)	4	5	2	2	2	3	1	2	<b>2.6</b>
CPU-D(s)	552	1523	579	1212	236	853	236	412	<b>700</b>
$\bar{\Delta}$ (%)	19	1	7	21	18	1	103	23	<b>23.5</b>

Table 4: CCFRLM-D vs. CCFRLM-B: Average CPU(s) and relative difference of actual coverage (%) on randomly generated instances

Results from Table 4 first show that the PSA approach (model CCFRLM-D) performs significantly better than the Bonferroni’s approximation (model CCFRLM-B) at providing good quality solutions for problem CCFRLM-EC. This can be seen by the fact that the average value of  $\bar{\Delta}$ , over the 80 considered test cases, is 23.5%. This means that the actual demand coverage provided by the recharging infrastructure is increased on average by 23.5% when using the PSA approximation of the joint chance constraints instead of the Bonferroni’s approximation.

This is mainly explained by the fact that the Bonferroni’s approximation relies on a conservative (pessimistic) approximation of the joint probability involved in constraints (8) so that model CCFRLM-B tends to underestimate the actual coverage probability of a trip. This overestimation of the risk of running out of fuel negatively impacts the performance of the location decisions made in model CCFRLM-B in two ways. Firstly, to cover a given trip, CCFRLM-B needs to open more stations than CCFRLM-D. As a consequence, in the optimal solution of model CCFRLM-B, the opened stations tend to be placed so as to cover a limited number of trips (most often those corresponding to the largest flows). As model CCFRLM-D is capable of covering the same set of trips with a smaller number of stations, it has the possibility to cover more trips with the remaining stations and to improve the global demand coverage. Secondly, when using Bonferroni’s approximation, some trips might seem “uncoverable” even if stations are opened at all nodes whereas they could in fact be covered provided some stations are opened at appropriately chosen locations. This is especially the case for trips which include one or several long arcs, i.e. arcs requiring an expected amount of energy close to the expected amount of energy available after recharging  $\bar{E}$ . For such trips, the sum of the risks of running out of fuel over the cycle segments visited when traveling in the outward and return directions might often exceed the maximum risk level  $\alpha$ . As constraints (12) of model CCFRLM-B uses this sum to approximate the risk of running out of fuel on a trip, a trip might seem “uncoverable” even if it is in fact coverable. In contrast, CCFRLM-D considers in constraints (25) the maximum expected risk over the cycle segments used on the path. Thus, even if this maximum expected risk is very close to  $\alpha$ , as long as it stays below  $\alpha$ , the trip is considered coverable by CCFRLM-D.

However, as shown in Tables 10 to 17, the benefit from using CCFRLM-D might be highly variable. For example, for set 80.40.5.20, we note that  $\bar{\Delta}$  varies between 0% and 943% depending on the network configuration (see line 3 of Table 16). This high variability is explained mainly by the impact of the road network structure on the performance of CCFRLM-B. As explained in the previous paragraph, a trip which includes one or several long arcs might seem “uncoverable” for CCFRLM-B whereas it could in fact be covered in practice. Therefore, if such a trip corresponds to a large EV flow, the impact on the actual coverage difference between model CCFRLM-B and model CCFRLM-D can be significant (see e.g. test cases 80.40.5.20 TC4, 80.40.25.20 TC4 and 100.50.25.20 TC3). Moreover, if there are several long trips with similar large EV flows in the network, model CCFRLM-B will tend to place stations so as to cover a limited number of these trips, thus leaving trips with large flows uncovered. In contrast, model CCFRLM-D will be able to share out the available stations among all these trips and to obtain a better demand coverage (e.g. test case 100.50.5.10 TC1). On the contrary, if the largest EV flows in the network correspond to a limited number of short trips, model CCFRLM-B will be able to find a solution in which all these flows are covered using the allowed number of stations. For such a network configuration, the actual coverage difference between the



two models is close to 0 so that the Bonferroni approach can be considered as a good approximation of the original stochastic problem (see e.g. test cases 100.50.5.10 TC3 and 100.50.25.10 TC3). Finally, we note that there are two test cases (80.40.25.10 TC5 and 80.40.25.10 TC9) for which  $\bar{\Delta}$  has a negative value. The lower quality of the solution provided by CCFRLM-D in these two cases is explained by the fact that in some instances (2 out of 5 for TC5 and 1 out of 5 for TC9), the evaluation of the EV flow coverage obtained by using the sampled scenarios was not accurate enough so that poor station location decisions were made by the model. This might be improved by increasing the sample size, i.e. by using more than 500 scenarios to represent the stochastic power consumption on each arc. However, this issue happened in only 3 instances out of 400. Moreover, in general, the variability of the objective function of CCFRLM-D over the 5 scenario samples used in each test case seems acceptable. The coefficient of variation of  $\hat{Z}_D$  over the 5 corresponding instances is namely on average 0.7% for the test cases 100.50.X.X and 0.4 % for the test cases 80.40.X.X. We thus did not increase the size of the sample used to represent the stochastic power consumption in the PSA approach in order to avoid longer computation times.

The average CPU time in seconds, CPU-B (resp. CPU-D), required to solve model CCFRLM-B (resp. CCFRLM-D) to optimality is reported in Table 4 for each aggregated set of instances and in Tables 10 to 17 for each individual test case. As can be seen from the results displayed in these tables, the partial sample approximation leads to computation times significantly larger than the ones needed for the Bonferroni approximation. However, they remain on average below 20 minutes. Moreover, as previously discussed, the solution provided by the Bonferroni approximation can be of very poor quality in some cases. As determining the location of EV charging stations is a strategic long-term problem, increasing the computation time by a reasonable amount to obtain better quality solutions seems a relevant trade-off.

### 6.3.2 Impact of problem parameters variation

In our computational experiments, we varied the main parameters of the problem, namely the problem dimensions, the number of charging stations to be opened and the standard deviation of the stochastic parameters, in order to study their impact on the solution quality and on the computation time.

Tables 10 to 17 show that in 27 cases out of 40, a larger number of stations (i.e. increasing  $p$  from 5 to 25) leads to the same or to a smaller value of  $\bar{\Delta}$ . This seems logical since allowing a larger number of stations provides CCFRLM-B with more flexibility to cover the largest flows of the network. However, there are 13 cases out of 40 where the difference in actual coverage increases. This happens in particular when the network displays a large number of long and risky paths (see e.g. test cases 100.50.5.20 and 100.50.25.20). In this situation, the largest flows are difficult to cover with 5 stations even for CCFRLM-D, so that when the number of stations allowed is increased to 25, the largest flows become covered by CCFRLM-D while some of them still cannot be covered by CCFRLM-B, which leads to increasing the actual coverage difference.

Moreover, the variability of the random parameters also seem to have an impact on the solution quality. We namely observe in Tables 10 to 17 that for 27 test cases out of 40, a higher standard deviation (i.e. increasing  $\varphi$  from 10% to 20%) leads to an equal or larger value of  $\bar{\Delta}$ . This could be explained by the fact that with a higher variability of the battery energy status and power consumption, the risk of running out of fuel on a cycle segment, i.e. the value of coefficients  $g_{kl}^q$  in constraints (12) and of coefficients  $\gamma_{kl}^{qs}$  in constraints (25), tends to increase. As a consequence, the difference between the sum of the risk over all cycle segments of a trip and its maximum value will tend to increase. For both models, it will thus be more difficult to cover the demand with the available stations when  $\varphi$  increases but the increase in difficulty will be more acute for model CCFRLM-B than for model CCFRLM-D.

Results from Tables 10 to 17 also show that the average CPU time tends to increase with the value of  $p$ , which might be explained by the fact that the number of possible combinations for stations deployment increases with  $p$  as long as  $p \leq |\mathcal{N}|/2$ . Computation times also increase with the problem dimension. Thus, the average CPU time needed to solve CCFRLM-D increases from 434s to 966s when increasing the problem dimension from 80 nodes and 40 O-D nodes to 100 nodes and 50 O-D nodes.

## 6.4 Results for real-life networks

Section 6.3 was based on randomly generated instances and focused on comparing the performance of the partial sample approximation approach with the one of the Bonferroni's approximation at solving the chance constrained flow refueling location problem under battery energy status and power consumption uncertainties. In the present section, tests are based on instances using real-life networks reported in the literature: the California road network [31] and the Florida road network [16].

For these instances, we first seek to evaluate the impact on the solution quality of using simplifying modeling assumptions such as a symmetric power consumption and deterministic input data. We thus consider in Subsection 6.4.1 the deterministic model (2)-(6) and assess the value of taking into account an asymmetric power consumption on each arc of the network. We then estimate the value of the stochastic solution, i.e. the value of using a more complex stochastic model (in which uncertainties on the battery energy status and power consumption are explicitly taken into account) instead of a simpler deterministic counterpart (in which only the expected value of these uncertain parameters is considered). Finally, we compare in Subsection 6.4.3 the partial sample approximation with the Bonferroni’s approximation.

#### 6.4.1 Impact of considering an asymmetric power consumption

As stated in Section 1, contrary to most previously published works, we allow in our problem modeling the power consumption on a road segment to depend on the crossing direction. This might improve the practical relevancy of the model as, in real life, the terrain configuration, road gradient and traffic condition may differ on a road segment according to the crossing direction. However, relaxing the assumption of a symmetric power consumption significantly increases the size of the optimization problem as the number of assignment variables  $w_{kl}^q$  and range constraints (4) is doubled. It is thus important to understand the impact of relaxing this assumption on the quality of the obtained charging infrastructure.

To that end, we focus on the deterministic flow refueling location model and seek to evaluate the benefit  $\Delta^{asym}$  from explicitly taking into account an asymmetric power consumption. This benefit is computed as follows. For each considered instance:

- We build a symmetric power consumption matrix  $\check{t}$  in which the power consumption on arc  $(a, b)$ ,  $\check{t}_{ab}$ , is defined by  $\check{t}_{ab} = \check{t}_{ba} = \max(\bar{t}_{ab}, \bar{t}_{ba})$ . In  $\check{t}$ , the power consumption on a given arc of the network is thus set to the maximum consumption among the one needed to cross the arc in the outward direction and the one needed to cross it in the return direction.
- We solve the deterministic problem (2)-(6) using the symmetric power consumption matrix  $\check{t}$  and record the obtained station location decision  $\check{x}_{det}^{sym}$ .
- We compute  $\check{Z}_{det}^{sym}$ , the demand coverage obtained by solving the deterministic problem (2)-(6) with the original asymmetric deterministic power consumption matrix  $\bar{t}$  while imposing the stations to be located as described by  $\check{x}_{det}^{sym}$ .  $\check{Z}_{det}^{sym}$  can be understood as an evaluation of the actual demand coverage that would be obtained when using the charging infrastructure  $\check{x}_{det}^{sym}$  when the power consumption on each arc is in fact asymmetric.
- We compute  $Z_{det}^{asym}$  the demand coverage obtained by solving the deterministic problem (2)-(6) with the original asymmetric deterministic power consumption matrix  $\bar{t}$ .
- The benefit from using an asymmetric power consumption is evaluated as  $\Delta^{asym} = \frac{Z_{det}^{asym} - \check{Z}_{det}^{sym}}{\check{Z}_{det}^{sym}}$ .

Table 5 shows the results obtained on the Florida and California instances with three different values of the number of stations  $p$ : 5, 15 and 25. The average value of  $\Delta^{asym}$  over all considered instances is equal to 1.82 %, which represents a rather large amount of additional EV flows that can be covered. We note however that the value of  $\Delta^{asym}$  depends on the network structure and on the number of stations allowed.  $\Delta^{asym}$  is indeed higher for the California instance in which the largest flows are concentrated on the longest trips and for the cases corresponding to a small value of  $p$ .

Test instance	CA.5	CA.15	CA.25	FL.5	FL.15	FL.25	<b>Avg.</b>
$\Delta^{asym}(\%)$	6.75	2.02	0.02	1.19	0.33	0.59	<b>1.82</b>

Table 5: Benefit from considering asymmetric consumption evaluated on California and Florida networks for  $p \in \{5, 15, 25\}$

#### 6.4.2 Value of the stochastic solution

The deterministic variant of the FRLP, i.e. problem (2)-(6), is significantly easier to solve than most of its stochastic variants. However, in practice, many input parameters of the problem are subject to uncertainty. Namely, the power consumption on a road segment depends on the driving conditions, which are subject to large and unpredictable

variations. Similarly, the energy available in the battery when an EV leaves a charging station depends on variable factors such as the age and technology of the battery and the time spent at the charging station and is thus also a stochastic input parameter. Neglecting these uncertainties as is done in the deterministic FRLP might lead to sub-optimal location decisions, i.e. to a charging infrastructure which, when implemented in a stochastic environment, will provide a smaller charging demand coverage than the one that could have been obtained. In order to assess the value of using a more complex stochastic model, we consider two measures:  $\Delta_D^{stoch}$ , the value of the stochastic solution provided by CCFRLM-D, and  $\Delta_B^{stoch}$ , the value of the stochastic solution provided by CCFRLM-B.

These two measures are computed by first estimating  $\hat{Z}_{det}^{asym}$ , the actual demand coverage that would be provided by the optimal solution of the deterministic problem (2)-(6) with the asymmetric power consumption  $\bar{t}$  if it were implemented in a stochastic environment. Following the procedure described in Section 6.2,  $\hat{Z}_{det}^{asym}$  is obtained by evaluating the performance of the deterministic location solution under a large number  $S'$  of scenarios for the battery energy status after recharging and the power consumption on each arc.  $\Delta_D^{stoch}$  and  $\Delta_B^{stoch}$  can then be defined by:  $\Delta_D^{stoch} = \frac{\hat{Z}_D - \hat{Z}_{det}^{asym}}{\hat{Z}_{det}^{asym}}$  and  $\Delta_B^{stoch} = \frac{\hat{Z}_B - \hat{Z}_{det}^{asym}}{\hat{Z}_{det}^{asym}}$  where  $\hat{Z}_D$  (resp.  $\hat{Z}_B$ ) corresponds to the estimated actual demand coverage provided by the optimal solution of CCFRLM-D (resp. CCFRLM-B).

Tables 6 and 7 show the results obtained on the California and Florida network instances with two possible values of the number of stations  $p$  (5 and 25) and two possible values of the relative standard deviation  $\varphi$  (10% and 20%). We note that over all considered instances, the average value of  $\Delta_D^{stoch}$  (resp.  $\Delta_B^{stoch}$ ) is 16.7% (resp. 13.5%). This means that by using a stochastic model, we are able to increase the actual demand coverage by an average of 16.7% (resp. 13.5%) when the partial sample approximation (resp. the Bonferroni's approximation) is used to handle the joint chance-constraints. This clearly shows the practical interest of using a stochastic model instead of a simpler deterministic counterpart.

However, there are several points worth raising here. Firstly, one can notice that the actual demand coverage provided by the solution of CCFRLM-B is sometimes less than or equal to the one provided by the deterministic solution (see cases *FL.5.10* and *FL.5.20*). As already discussed in Section 6.3 for the case of randomly generated networks, this might be explained by the fact that model CCFRLM-B sometimes largely overestimates the risk of running out of fuel between two stations and thus makes poor location decisions. Second, we observe that for cases *CA.5.10* and *CA.5.20*, the values of  $\Delta_B^{stoch}$  and  $\Delta_D^{stoch}$  cannot be computed because  $\hat{Z}_{det}^{asym}$  is equal to 0. This means that the charging infrastructure obtained by the deterministic model does not allow to cover any flow when evaluated by the post-optimization simulation procedure. This is explained by the structure of the California network in which the largest flows are concentrated on the longest trips. The location decisions of the deterministic model thus aim at covering these longest trips while using only 5 stations. This results in an infrastructure in which the stations are spaced from one another at a distance close to the expected battery energy status after recharging  $\bar{E}$ . When implemented in a stochastic environment, such an infrastructure is not capable to cover any flow as the joint probability of running out of charge when crossing the long segments between two opened stations is above the maximum acceptable risk level.

Test instance	<i>CA.5.10</i>	<i>CA.25.10</i>	<i>CA.5.20</i>	<i>CA.25.20</i>	<b>Avg.</b>
$\Delta_B^{stoch}$ (%)	-	12.3	-	24.7	<b>18.5</b>
$\Delta_D^{stoch}$ (%)	-	15.4	-	29.7	<b>22.5</b>

Table 6: Value of using a stochastic model evaluated on California network

Test instance	<i>FL.5.10</i>	<i>FL.25.10</i>	<i>FL.5.20</i>	<i>FL.25.20</i>	<b>Avg.</b>
$\Delta_B^{stoch}$ (%)	0	6.9	-0.9	27.9	<b>8.5</b>
$\Delta_D^{stoch}$ (%)	0.9	6.9	6.1	29.8	<b>10.9</b>

Table 7: Value of using a stochastic model evaluated on Florida network

### 6.4.3 Comparative analysis of the solution approaches

In order to confirm the conclusions obtained in Section 6.3 using randomly generated instances, we propose here to test the partial sample approximation and the Bonferroni's approaches on California and Florida road networks. The value of  $\bar{\Delta}$ , the relative difference between the actual EV flow coverage obtained using each of the two solution methods, as well as the average computation times are summarized in Table 8 (resp. in Table 9) for all test cases corresponding to the California network (resp. to the Florida network).

Test instance	<i>CA.5.10</i>	<i>CA.25.10</i>	<i>CA.5.20</i>	<i>CA.25.20</i>	<b>Avg.</b>
CPU-B(s)	399	1693	2652	36000	<b>10186</b>
CPU-D(s)	8357	22635	10758	36000	<b>19438</b>
$\bar{\Delta}(\%)$	2	3	1	4	<b>2.5</b>

Table 8: CCFRLM-D vs. CCFRLM-B: Average CPU(s) and relative difference of actual coverage (%) on California real-life instance

Test instance	<i>FL.5.10</i>	<i>FL.25.10</i>	<i>FL.5.20</i>	<i>FL.25.20</i>	<b>Avg.</b>
CPU-B(s)	165	380	26	50	<b>155</b>
CPU-D(s)	6857	10967	6151	12978	<b>9238</b>
$\bar{\Delta}(\%)$	1	0	7	1	<b>2.2</b>

Table 9: CCFRLM-D vs. CCFRLM-B: Average CPU(s) and relative difference of actual coverage (%) on Florida real-life instance

These results show that the proposed PSA approach also outperforms the Bonferroni approximation on-real life networks in terms of solution quality. Namely, on average, the actual demand coverage is increased by 2.35% when using the PSA approach rather than the Bonferroni approximation. Even if this improvement is smaller than the one observed on randomly generated instances, the gain in terms of additional EV flows covered is still significant enough to motivate the use of the PSA approach.

We also observe that the computation times needed to solve both models are longer for the real-life instances than for the randomly generated ones. This is an expected outcome since real-life networks are of larger dimensions. Moreover, computation times remain longer for model CCFRLM-D than for model CCFRLM-B. In particular, within the computation time limit of 10 hours, model CCFRLM-D could not find guaranteed optimal solutions for 3 out of 40 instances (the optimality gap lied within 0.1%-5%) while this happened in 1 out of 8 instances for model CCFRLM-B (the optimality gap was 0.3%). Moreover, for 3 out of 40 instances, model CCFRLM-D was not able to find any feasible solution within 10h of calculation: the corresponding instances were not considered in the results reported in Tables 8 and 9.

## 7 Conclusion and future research

We studied the flow refueling location problem under a stochastic driving range and focused on relaxing several modeling assumptions previously used in the literature to handle this problem. We first considered an asymmetric power consumption over the arcs of the network. Namely, due to terrain configuration and road gradient, the power consumption on a road segment might depend on the crossing direction. We then took into account uncertainties related to the energy available in the battery after recharging at a station, which are faced by all EV drivers who travel through the road network, as well as uncertainties related to the power consumption on each portion of the road network, which result mainly from the local traffic conditions. Moreover, in order to enhance the practical relevancy of the model and take into account the dependencies in the traffic conditions in different areas of the road network, we considered correlations between the stochastic power consumption on different arcs of the network. We developed a chance constrained flow refueling location model in which a flow is considered covered if its coverage probability remains above a predefined minimum value and the objective is to maximize the total flow covered.

To solve the resulting stochastic problem, we compared two solution approaches: the Bonferroni approximation and the Partial Sample Approximation (PSA) approach. Both methods allow to overcome the difficulty of calculating the joint probability involved in the chance constraints and lead to the formation of a mixed-integer linear program providing an approximate solution of the original stochastic problem. We carried out numerical experiments on a set of medium-size randomly generated and real life instances. Our results show that the proposed PSA approach outperforms the Bonferroni approach in terms of solution quality and gives station locations which provide a significantly improved demand coverage in practice.

However, the computation times needed to solve the MILP resulting from the PSA approach are longer than those needed to solve the MILP resulting from the Bonferroni approach. An interesting direction for future research would thus be to study solution methods such as the Benders decomposition in order to improve the solution time of the PSA approach. On a longer perspective, introducing realistic features in the FRLP with a stochastic driving range such as a limited capacity of the charging stations or deviations from the shortest path would also be relevant.

## A Computation of the probability distributions

In Section 5,  $g_{kl}^q$  was defined for a segment  $[k, l]$  of trip  $q$  as the probability that a vehicle recharging at a station in node  $k$  runs out of fuel before reaching node  $l$ , i.e.  $g_{kl}^q = P(\widetilde{t}_{kl}^q \geq \widetilde{E})$ .

In order to evaluate  $g_{kl}^q$ , we define  $\widetilde{\theta}_{kl}^q = \widetilde{t}_{kl}^q - \overline{t}_{kl}^q$ , with  $\overline{t}_{kl}^q$  the expected length of segment  $[k, l]$  on trip  $q$ . Let  $\widetilde{E}_{kl}^q = \widetilde{E} - \widetilde{\theta}_{kl}^q$ , accordingly, we have:  $P(\widetilde{t}_{kl}^q \geq \widetilde{E}) = P(\widetilde{E} - \widetilde{\theta}_{kl}^q \leq \overline{t}_{kl}^q) = P(\widetilde{E}_{kl}^q \leq \overline{t}_{kl}^q)$ ,  $\widetilde{E}_{kl}^q$  is a random variable that can be interpreted as the energy available in the battery for crossing segment  $[k, l]$  of trip  $q$ . Assuming that  $G_{\widetilde{E}_{kl}^q}$  denotes the cumulative density function of random variable  $\widetilde{E}_{kl}^q$ , we have  $g_{kl}^q = G_{\widetilde{E}_{kl}^q}(\overline{t}_{kl}^q)$ . Hence, knowing the probability distribution of the random variable  $\widetilde{E}_{kl}^q$ , it is possible to evaluate parameter  $g_{kl}^q$ .

We assume in what follows that the random battery energy status after recharging  $\widetilde{E}$  follows a normal distribution:  $\widetilde{E} \hookrightarrow \mathcal{N}(\overline{E}, \sigma_E^2)$ . We also assume that random power consumption  $\widetilde{t}_{ab}$  on arcs  $(a, b) \in \mathcal{A}$  follows a multivariate normal distribution characterized by an expected consumption vector  $\overline{\mathbf{t}}$  and a covariance matrix  $\sigma$ :

$$\overline{\mathbf{t}} = \begin{pmatrix} \dots \\ \overline{t}_{ab} \\ \dots \\ \overline{t}_{cd} \\ \dots \end{pmatrix} \text{ and } \sigma = \begin{pmatrix} \sigma_{ab}^2 & \dots & \dots & \sigma_{ab,cd} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \sigma_{ab,cd} & \dots & \dots & \sigma_{cd}^2 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

We denote  $\overline{t}_{ab}$  the expectation of the random consumption  $\widetilde{t}_{ab}$  on arc  $(a, b) \in \mathcal{A}$ ,  $\sigma_{ab}$  the standard deviation of  $\widetilde{t}_{ab}$  and  $\sigma_{ab,cd}$  the covariance between consumption  $\widetilde{t}_{ab}$  on arc  $(a, b) \in \mathcal{A}$  and consumption  $\widetilde{t}_{cd}$  on arc  $(c, d) \in \mathcal{A}$ .

Let us consider the example of a cycle segment  $[k, l]$  of trip  $q$  composed of three arcs  $(k, d)$ ,  $(d, e)$  and  $(e, l)$  (see Figure 1). According to the definition of cycle segment presented in Section 3, the random consumption  $\widetilde{t}_{kl}^q$  on segment  $[k, l]$  of trip  $q$  follows a normal distribution and can be evaluated as follows:

- Outward direction trip: If  $k = O^q$  (respectively  $l = D^q$ ) and  $l \neq D^q$  (respectively  $k \neq O^q$ ),  $\widetilde{t}_{kl}^q = \widetilde{t}_{kd} + \widetilde{t}_{de} + \widetilde{t}_{el} + \widetilde{t}_{le} + \widetilde{t}_{ed} + \widetilde{t}_{dk}$  since in this case the round trip on segment  $[k, l]$  has to be refueled with a station located at node  $l$  (respectively  $k$ ) or after.
- Return direction trip: If  $k = D^q$  (respectively  $l = O^q$ ) and  $l \neq O^q$  (respectively  $k \neq D^q$ ),  $\widetilde{t}_{kl}^q = \widetilde{t}_{kd} + \widetilde{t}_{de} + \widetilde{t}_{el} + \widetilde{t}_{le} + \widetilde{t}_{ed} + \widetilde{t}_{dk}$  since in this case the round trip on segment  $[k, l]$  has to be refueled with a station located at node  $l$  (respectively  $k$ ) or before.
- If  $k \notin \{O^q, D^q\}$  and  $l \notin \{O^q, D^q\}$ ,  $\widetilde{t}_{kl}^q = \widetilde{t}_{kd} + \widetilde{t}_{de} + \widetilde{t}_{el}$

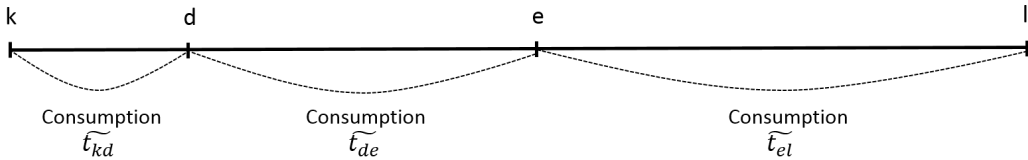


Figure 1: Example of a cycle segment  $[k, l]$

The average consumption  $\overline{t}_{kl}^q$  and the variance  $\sigma_{kl}^2$  on segment  $[k, l]$  are thus calculated as follows:

- Outward direction trip: If  $k = O^q$  (respectively  $l = D^q$ ) and  $l \neq D^q$  (respectively  $k \neq O^q$ )

$$\overline{t}_{kl}^q = \sum_{(a,b) \in \mathcal{A}_{kl}^q \cup \mathcal{A}_{lk}^q} \overline{t}_{ab}$$

$$\sigma_{kl}^2 = \sum_{(a,b) \in \mathcal{A}_{kl}^q \cup \mathcal{A}_{lk}^q} \sigma_{ab}^2 + 2 \sum_{\substack{(a,b) \in \mathcal{A}_{kl}^q \cup \mathcal{A}_{lk}^q \\ (c,d) \in \mathcal{A}_{kl}^q \cup \mathcal{A}_{lk}^q \\ o(a,b) < o(c,d)}} \sigma_{ab,cd}$$

We denote  $o(a, b)$  the order of visiting arc  $(a, b)$  on the round trip from  $k$  to  $l$ .

- Return direction trip: If  $k = D^q$  (respectively  $l = O^q$ ) and  $l \neq O^q$  (respectively  $k \neq D^q$ )

$$\begin{aligned}\overline{t_{kl}^q} &= \sum_{(a,b) \in \mathcal{A}_{kl}^q \cup \mathcal{A}_{lk}^q} \overline{t_{ab}} \\ \sigma_{kl}^2 &= \sum_{(a,b) \in \mathcal{A}_{kl}^q \cup \mathcal{A}_{lk}^q} \sigma_{ab}^2 + 2 \sum_{\substack{(a,b) \in \mathcal{A}_{kl}^q \cup \mathcal{A}_{lk}^q \\ (c,d) \in \mathcal{A}_{kl}^q \cup \mathcal{A}_{lk}^q \\ o(a,b) < o(c,d)}} \sigma_{ab,cd}\end{aligned}$$

- If  $k \notin \{O^q, D^q\}$  and  $l \notin \{O^q, D^q\}$

$$\begin{aligned}\overline{t_{kl}^q} &= \sum_{(a,b) \in \mathcal{A}_{kl}^q} \overline{t_{ab}} \\ \sigma_{kl}^2 &= \sum_{(a,b) \in \mathcal{A}_{kl}^q} \sigma_{ab}^2 + 2 \sum_{\substack{(a,b) \in \mathcal{A}_{kl}^q \\ (c,d) \in \mathcal{A}_{kl}^q \\ o(a,b) < o(c,d)}} \sigma_{ab,cd}\end{aligned}$$

Hence, we can determine the probability distribution of random variable  $\widetilde{E}_{kl}^q$  required to calculate parameter  $g_{kl}^q$ . As earlier explained,  $\widetilde{E}_{kl}^q = \widetilde{E} - \widetilde{\theta}_{kl}^q = \widetilde{E} - \widetilde{t}_{kl}^q + \overline{t_{kl}^q}$ . According to the assumption made in subsection 5.2, random variables  $\widetilde{E}$  and  $\widetilde{t}_{kl}^q$  are independent, therefore,  $\widetilde{E}_{kl}^q \hookrightarrow \mathcal{N}(\overline{E}, \sigma_E^2 + \sigma_{kl}^2)$ .

## B Numerical results for each randomly generated network instance

Test case	TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	Avg
CPU-B(s)	7	5	4	3	3	3	5	4	3	4	<b>4</b>
CPU-D(s)	672	575	589	468	391	516	661	575	508	569	<b>552</b>
$\overline{\Delta}(\%)$	130	35	0	0	0	18	1	9	0	0	<b>19</b>

Table 10: CCFRLM-D vs. CCFRLM-B: Average CPU(s) and relative difference of actual coverage (%) for set 100.50.5.10

Test case	TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	Avg
CPU-B(s)	5	11	3	6	3	5	5	4	3	5	<b>5</b>
CPU-D(s)	1586	1170	2805	613	1311	1070	1635	1064	1219	2758	<b>1523</b>
$\overline{\Delta}(\%)$	3	0	0	1	0	3	1	2	1	1	<b>1</b>

Table 11: CCFRLM-D vs. CCFRLM-B: Average CPU(s) and relative difference of actual coverage (%) for set 100.50.25.10

Test case	TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	Avg
CPU-B(s)	2	2	2	2	2	1	2	1	2	2	<b>2</b>
CPU-D(s)	649	675	471	466	456	635	550	666	630	596	<b>579</b>
$\overline{\Delta}(\%)$	5	5	33	12	0	1	0	2	12	0	<b>7</b>

Table 12: CCFRLM-D vs. CCFRLM-B: Average CPU(s) and relative difference of actual coverage (%) for set 100.50.5.20

Test case	TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	Avg
CPU-B(s)	3	2	2	2	2	3	2	2	3	3	<b>2</b>
CPU-D(s)	1556	1443	789	849	1044	693	825	693	1038	3191	<b>1212</b>
$\overline{\Delta}(\%)$	1	1	103	13	58	5	3	2	21	1	<b>21</b>

Table 13: CCFRLM-D vs. CCFRLM-B: Average CPU(s) and relative difference of actual coverage (%) for set 100.50.25.20

Test case	TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	Avg
CPU-B(s)	1	2	2	2	2	2	1	2	2	1	<b>2</b>
CPU-D(s)	199	190	205	235	269	285	241	226	276	236	<b>236</b>
$\bar{\Delta}(\%)$	0	0	57	23	17	2	12	42	3	23	<b>18</b>

Table 14: CCFRLM-D vs. CCFRLM-B: Average CPU(s) and relative difference of actual coverage (%) for set 80.40.5.10

Test case	TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	Avg
CPU-B(s)	2	2	2	5	2	3	2	2	3	2	<b>3</b>
CPU-D(s)	369	233	260	1529	495	1307	336	358	3037	607	<b>853</b>
$\bar{\Delta}(\%)$	0	0	3	1	-3	0	6	0	-4	4	<b>1</b>

Table 15: CCFRLM-D vs. CCFRLM-B: Average CPU(s) and relative difference of actual coverage (%) for set 80.40.25.10

Test case	TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	Avg
CPU-B(s)	1	1	2	1	2	1	1	1	1	1	<b>1</b>
CPU-D(s)	241	203	214	229	242	271	208	249	257	249	<b>236</b>
$\bar{\Delta}(\%)$	2	0	51	943	0	20	2	0	8	9	<b>103</b>

Table 16: CCFRLM-D vs. CCFRLM-B: Average CPU(s) and relative difference of actual coverage (%) for set 80.40.5.20

Test case	TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	Avg
CPU-B(s)	2	2	1	2	1	2	2	2	2	1	<b>2</b>
CPU-D(s)	318	266	338	326	323	1108	295	275	621	252	<b>412</b>
$\bar{\Delta}(\%)$	1	0	7	181	0	27	3	1	6	2	<b>23</b>

Table 17: CCFRLM-D vs. CCFRLM-B: Average CPU(s) and relative difference of actual coverage (%) for set 80.40.25.20

## References

- [1] C. Bonferroni. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8:3–62, 1936.
- [2] G. Calafiore and M. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102:25–46, 2005.
- [3] I. Capar, M. Kuby, V. J. Leon, and Y.-J. Tsai. An arc cover–path-cover formulation and strategic analysis of alternative-fuel station locations. *European Journal of Operational Research*, 227:142–151, 2013.
- [4] J. Cheng, C. Gicquel, and A. Lisser. Partial sample approximation method for chance-constrained problems. *Optimization Letters*, 13(4):657–672, 2019.
- [5] S. Davidov and M. Pantos. Impact of stochastic driving range on the optimal charging infrastructure expansion planning. *Energy*, 141:603–612, 2017.
- [6] H. De Vries and E. Duijzer. Incorporating driving range variability in network design for refueling facilities. *Omega*, 69:102–114, 2017.
- [7] A. Fotheringham and M. O’Kelly. Spatial interaction models: Formulations and applications. *Kluwer Academic Publishers*, 1989.
- [8] I. Frade, A. Ribeiro, G. Goncalves, and A. Pais Antunes. Optimal locations of charging stations for electric vehicles in a neighborhood in Lisbon, Portugal. *Transportation Research Record: Journal of The Transportation Research Board*, 2252:91–98, 2011.
- [9] M. J. Hodgson. A flow capturing location-allocation model. *Geographical analysis*, 22:270–279, 1990.
- [10] M. Hosseini and S. A. MirHassani. Refueling station location problem under uncertainty. *Transportation Research Part E*, 84:101–116, 2015.

- [11] W. Jing, K. An, and M. Ramezani. Location design of electric vehicle charging facilities: A path-distance constrained stochastic user equilibrium approach. *Journal of advanced transportation*, 2017.
- [12] J. Jung, J. Y. J. Chow, R. Jayakrishnan, and J. Y. Park. Stochastic dynamic itinerary interception refueling location problem with queue delay for electric taxi charging stations. *Transportation Research Part C*, 40:123–142, 2014.
- [13] M. Kchaou Boujelben and C. Gicquel. Efficient solution approaches for locating electric vehicle fast charging stations under driving range uncertainty. *Computers and Operations Research*, 109:288–299, 2019.
- [14] J. G. Kim and M. Kuby. A network transformation heuristic approach for the deviation flow refueling location model. *Computers and Operations Research*, 40:1122–1131, 2013.
- [15] M. Kuby and S. Lim. The flow-refueling location problem for alternative-fuel vehicles. *Socio-Economic Planning Sciences*, 39:125–145, 2005.
- [16] M. Kuby, L. Lines, R. Schultz, Z. Xie, J. G. Kim, and S. Lim. Optimization of hydrogen stations in Florida using the flow-refueling location model. *International Journal Of Hydrogen Energy*, 34:6045–6064, 2009.
- [17] C. Lee and J. Han. Benders-and-price approach for electric vehicle charging station location problem under probabilistic travel range. *Transportation Research Part B*, 106:130–152, 2017.
- [18] J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal of Optimization*, 19(2):674–699, 2008.
- [19] I. Y. Mak, Y. Rong, and Z. J. M. Shen. Infrastructure planning for electric vehicles with battery swapping. *Management Science*, 59:1557–1575, 2013.
- [20] S. Mehar and S. Senouci. An optimization location scheme for electric charging stations. *International Conference on Smart Communications in Network Technologies SaCoNet*, Paris, France, 2013.
- [21] S. A. MirHassani and R. Ebrazi. A flexible reformulation of the refueling station location problem. *Transportation Science*, 47(4):455–628, 2013.
- [22] A. Nemirovski and A. Shapiro. Scenario approximations of chance constraints. In *Probabilistic and Randomized Methods for Design Under Uncertainty*, pages 3–48. Springer, 2005.
- [23] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal of Optimization*, 17:969–996, 2006.
- [24] R. Riemann, D. Z. W. Wang, and F. Busch. Optimal location of wireless charging facilities for electric vehicles: Flow-capturing location model with stochastic user equilibrium. *Transportation Research Part C*, 58:1–12, 2015.
- [25] R. T. Rockafellar and S. P. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2000.
- [26] B. Rostami, G. Desaulniers, F. Errico, and A. Lodi. The vehicle routing problem with stochastic and correlated travel times. *Technical report, Data Science for real-time decision-making*, DS4DM-2017-016, 2017.
- [27] M. Schiffer and G. Walther. Strategic planning of electric logistics fleet networks: A robust location-routing approach. *Omega*, 80:31–42, 2018.
- [28] C. Upchurch and M. Kuby. A capacitated model for location of alternative-fuel stations. *Geographical analysis*, 41:89–106, 2009.
- [29] Y. W. Wang and C. C. Lin. Locating road-vehicle refueling stations. *Transportation Research Part E*, 45:821–829, 2009.
- [30] F. Wu and R. Sioshansi. A stochastic flow-capturing model to optimize the location of fast-charging stations with uncertain electric vehicle flows. *Transportation Research Part D*, 53:354–376, 2017.
- [31] B. Yildiz, O. Arslan, and O. Ekin Karasan. A branch and price approach for routing and refueling station location model. *European Journal of Operational Research*, 248:815–826, 2016.
- [32] B. Yildiz, E. Olcaytu, and A. Sen. The urban recharging infrastructure design problem with stochastic demands and capacitated charging stations. *Transportation Research Part B*, 119:22–44, 2019.