DEAR: Dynamic Electric Ambulance Redeployment

Lukas Rottkamp*
MCML, LMU Munich
Munich, Germany
rottkamp@cip.ifi.lmu.de

Niklas Strauß*
MCML, LMU Munich
Munich, Germany
strauss@dbs.ifi.lmu.de

Matthias Schubert
MCML, LMU Munich
Munich, Germany
schubert@dbs.ifi.lmu.de

ABSTRACT

Dynamic Ambulance Redeployment (DAR) is the task of dynamically assigning ambulances after incidents to base stations to minimize future response times. Though DAR has attracted considerable attention from the research community, existing solutions do not consider using electric ambulances despite the global shift towards electric mobility. In this paper, we are the first to examine the impact of electric ambulances and their required downtime for recharging to DAR and demonstrate that using policies for conventional vehicles can lead to a significant increase in either the number of required ambulances or in the response time to emergencies. Therefore, we propose a new redeployment policy that considers the remaining energy levels, the recharging stations' locations, and the required recharging time. Our new method is based on minimizing energy deficits (MED) and can provide well-performing redeployment decisions in the novel Dynamic Electric Ambulance Redeployment problem (DEAR). We evaluate MED on a simulation using real-world emergency data from the city of San Francisco and show that MED can provide the required service level without additional ambulances in most cases. For DEAR, MED outperforms various established state-of-the-art solutions for conventional DAR and straightforward solutions to this setting.

CCS CONCEPTS

• **Information systems → Spatial-temporal systems**; • **Computing methodologies → Simulation environments**.

KEYWORDS

Ambulance Redeployment, Optimization, Spatio-Temporal Data

ACM Reference Format:
Lukas Rottkamp, Niklas Strauß, and Matthias Schubert. 2023. DEAR: Dynamic Electric Ambulance Redeployment. In *Symposium on Spatial and Temporal Data (SSTD '23), August 23–25, 2023, Calgary, AB, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3609956.3609959

1 INTRODUCTION

The Emergency Medical Service (EMS) is a critical part of health infrastructure all over the world [15]. Paramedics are often the first professional aid in health emergencies and are responsible for safe and quick transport to a secondary care unit such as a hospital. A low response time to emergency calls has increased survival and recovery rates in life-threatening health conditions such as cardiac arrest [5, 18]. Ambulance response times to emergencies depends on various factors, such as the emergency call itself, the processing time needed for dispatch, the readiness of a qualified paramedic team, and its travel time to the incident location. Travel time is a substantial factor. While it can be accelerated by using high-powered vehicles and specialized training for driving in emergency conditions, the initial distance of the ambulance to the incident site is the most prominent factor, with various approaches trying to minimize this distance by proper ambulance placement.

Today, most ambulances are outfitted with internal combustion engines (ICE) using fossil fuels. However, the growing public demand for less air pollution and less release of greenhouse gases promotes the transition towards electric vehicles (EV). Electric ambulances further come with additional benefits, such as a smoother acceleration improving in-ambulance care. Thus, a first generation of electric ambulances is already commercially available.

Ambulances are usually positioned at base stations strategically placed over a city or coverage area to minimize incident response times. Incoming emergency calls are assigned to an ambulance, which drives to the incident location. Some incidents can be resolved on-site, while in other cases, patients need to be transported to a hospital. After completing their assignment, ambulances return to a base station. While ambulances could return to their origin station, it is often advisable to select another base station based on the actual ambulance distribution at this time. This selection of base stations is known as the Dynamic Ambulance Redeployment (DAR) problem in literature [13, 16, 23].

In this paper, we show that existing approaches do not perform well when confronted with electric ambulances. First, we present a formal definition of the Dynamic Electric Ambulance Redeployment Problem (DEAR), which extends existing DAR formalizations by battery levels, range restrictions, charging stations, and recharging. Based on this extension, we can examine the performance of established state-of-the-art methods for dynamic ambulance redeployment, which do not consider these aspects. Afterwards, we present the minimizing energy deficits (MED) approach, designed to avoid these shortcomings and provide state-of-the-art ambulance redeployment for E-Ambulances. Our method is based on matching the predicted future demand in the area of each base station to the joint energy level of the ambulances. The energy level of vehicles at a base station is extrapolated for the same time frame as the future demand and considers any recharging activity increasing the energy level. Based on both estimations on future development, MED assigns ambulances to those base stations where the deficits between the energy level and the demand are expected to be the largest. We compare MED to various state-of-the-art conventional

*Both authors contributed equally to this research.

SSTD '23, August 23–25, 2023, Calgary, AB, Canada

ACM ISBN 979-8-4007-0899-2/23/08.
https://doi.org/10.1145/3609956.3609959

ambulance redeployment methods on an extended environment of [23]. Our results demonstrate that the conventional DAR methods suffer significant performance decreases in various settings. In contrast, MED can cope well with the requirements of E-Ambulances, often compensating for their drawbacks against using conventional ICE ambulances.

To summarize, our contributions are as follows:

- We formalize DEAR, an extension of the DAR problem considering electric ambulances.
- We extended a DAR simulation environment based on real-world data to consider the DEAR setting and examine the performance of conventional DAR methods.
- We propose MED and present experimental results showing that it copes well with DEAR compared to existing DAR methods and basic DEAR approaches.

The remainder of this paper is structured as follows: Related work is presented in Section 2. We then formulate the Dynamic Electric Ambulance Redeployment Problem (DEAR) in Section 3 and propose MED in Section 4. We evaluate established DAR approaches and MED for DEAR using a simulation based on real-world incident data from San Francisco in Section 5 and summarize our work in Section 6.

2 RELATED WORK

The ambulance location problem (ALP) is an established research topic. Existing approaches can be classified into static and dynamic methods: In static methods, ambulances are stationed at fixed base stations and always return to the same base station after an incident has been handled [7, 8, 19]. One way to obtain a static assignment is to solve the *Maximum Expected Covering Location Problem* (MEXCLP) [8, 13]. Its solution maximizes the expected coverage of incident locations. In contrast to the *Maximum Coverage Location Problem* [7] it is based on, the underlying model assumes an ambulance to be *busy* with a certain probability. In this way, ambulances that are unavailable due to being on a mission, are not included in the coverage calculation. This reasonable modification has been proven to be advantageous compared to earlier methods [12, 13]. *Expected Response Time Model* (ERTM) [3] is another static approach that has shown excellent performance due to its direct minimization of the expected response time [3, 23].

Current state-of-the-art ALP solutions use a dynamic assignment due to the volatility of the problem [13]. The dynamic assignment of ambulances is also called *Real-Time Ambulance Redeployment Problem* or *Dynamic Ambulance Redeployment Problem* (DAR). Dynamic redeployment leads to better response times than static return policies because the stochastic nature of incoming emergency calls can lead to imbalances in ambulance distribution which are ignored by static approaches [10, 11]. The redeployment decision is primarily based on the locations of ambulances and base stations but may also take other factors, such as demand distributions, into account. The *DMEXCLP* approach by [13] is a dynamic variation of MEXCLP. At each redeployment step, it selects the base station providing the largest coverage increase in the respective situation according to the MEXCLP strategy. This way, DMEXCLP takes the actual distribution of ambulances into account. A reinforcement-learning based approach "Reinforcement Learning Deep Score Network" (DRLSN) is presented by [14].

A vision paper by [20] highlights the growing importance of electric ambulances and the associated challenge of keeping a fleet of ambulances charged. It suggests a high-level framework for ambulance scheduling concerning the optimal use of renewable energy sources, including predictive components for patient demand and energy production and use. Though this work is related, it does neither propose a formalization of DEAR nor does it provide a method for the redeployment problem for electric ambulances.

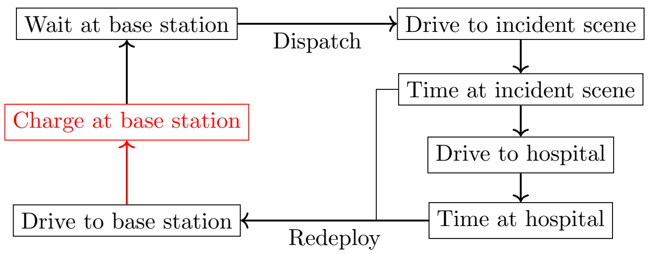


Figure 1: Simplified schematic overview of the modeled EMS process. Specifics for electric ambulances are shown in red.

3 PROBLEM DEFINITION

In this section, we will provide a formal definition of the Dynamic Electric Ambulance Redeployment (DEAR) problem, outlining the operational process of the Emergency Medical Services (EMS) provider and considering the specifics of electric ambulances. Figure 1 provides a visual representation of the EMS process. When an incident occurs, the EMS operator receives a call, and an available ambulance is dispatched from a base station to the incident location. In our scenario, the ambulance closest in driving time is dispatched to ensure a prompt response. If no ambulance is available, the incident is handled as soon as an ambulance becomes available again. Upon arrival at the incident, on-site care is provided to the patient. Depending on the patient's condition, subsequent transport to a hospital may be necessary. Otherwise, the ambulance is redeployed from the incident site to a base station. Once the ambulance arrives at a base station, it becomes idle and available for dispatch. Considering electric vehicles introduces unique challenges compared to Internal Combustion Engine (ICE) vehicles. The downtime for refueling ICE vehicles is typically not a significant concern due to their long ranges and fast refueling times. However, electric vehicles have shorter ranges and require substantial charging time. Therefore, factors such as charging downtime, battery levels, and the availability of fast chargers at base stations need to be considered in the EMS process. It is crucial only to dispatch an electric ambulance if its battery is sufficiently charged to not run out of energy while handling the incident. Therefore, we define a minimum dispatch range τ_{MDR} (measured in time units) as the worst-case trip, starting from the current base station to any incident location, followed by transportation to any hospital, and finally redeployment to a base station.

Electric ambulances can be charged at regular AC outlets (we refer to them as slow chargers), which are already available in large numbers at base stations. However, slow chargers have limited power output, resulting in extended charging times and longer downtimes of ambulances. Charging times can be significantly reduced by installing high-voltage DC chargers (fast chargers specifically installed for electric vehicles) at base stations. However, their number is limited because installation presents a significant cost factor and constraints caused by the capabilities of the energy grid.

Assigning chargers to ambulances at a base station requires a charging policy when the number of ambulances exceeds the number of fast chargers. The objective is to charge ambulances in a manner that allows them to reach the minimum dispatch range τ_{MDR} as quickly as possible, thereby maximizing the number of available ambulances. It is also important to avoid an unreasonably high number of re-plugging actions by staff. To achieve these goals, we implement the following approach: Ambulances below τ_{MDR} are categorized as high-priority and are charged first. If there are more high-priority ambulances than available chargers or fast chargers, the ambulance with the shortest time required to reach τ_{MDR} is prioritized for charging. This ensures that ambulances are prepared for service at the earliest possible time. Once an ambulance reaches τ_{MDR}, it becomes a low-priority ambulance. For charging low-priority ambulances, we prioritize ambulances with the lowest battery level to minimize the number of re-plugging actions. Replugging can occur when an ambulance at the station is sufficiently charged to provide the minimum dispatch range, is fully charged, arrives, or is dispatched.

Now, we present a formal definition of the novel DEAR problem, considering the aforementioned characteristics. In this task, an operator needs to dynamically select a base station to redeploy an ambulance to after the ambulance finishes handling an incident, either from the incident site or the hospital.

The road network is represented as a graph $G = (V, E)$, where V is the set of nodes representing locations in the road network, and E is the set of directed edges representing road segments connecting the nodes.

Incidents are emergencies requiring medical attention by an ambulance and are denoted as I. Each incident is mapped to the nearest node in the graph.

Base Stations Let W be the set of base stations available within the road network, where ambulances are stationed and dispatched to incidents. Each base station is mapped to the closest node in the road network. Base stations are equipped with charging infrastructure to support the operation of electric ambulances. They possess an unlimited number of slow chargers (regular AC outlets) and have varying numbers of fast chargers. Not all base stations are guaranteed to have fast chargers available.

Hospitals The set H represents the hospitals. Similar to base stations, hospitals are mapped to the closest node in the graph.

Ambulances are electric vehicles, introducing specific characteristics that affect their operational constraints. Key properties include battery level and capacity, energy use per time, and charging characteristics. The charging rate of an ambulance depends on various factors, including its current battery level and the power output of the charger. A linear charging function is utilized, although other charging functions may also be employed. We assume that all ambulances are the same type, i.e., their key properties are equal. Let us note that our method can easily be adapted to more specific settings if required. Ambulances are initially assigned to base stations and can be dynamically redeployed to other base stations depending on incident demand. We allow an ambulance to be redeployed only after finishing handling an incident.

Travel Times In our setting, the travel times $\tau(i, j)$ between two nodes $i, j \in V$ are assumed to be deterministic and do not vary with traffic conditions. When responding to an incident or transporting a patient, ambulances use lights and sirens to alert other drivers, enabling them to travel at fast speeds [4]. We denote the travel time with lights and sirens activated as $\hat{\tau}(i, j)$.

4 MED: MINIMIZE ENERGY DEFICIT

In this section, we introduce our approach Minimize Energy Deficit (MED) for the DEAR problem.

While approaches for solving the DAR problem can be applied, they do not take the additional complexity of electric ambulances into account. Our evaluation demonstrates that this leads to drastically degraded response times or requires multiple additional ambulances to maintain EMS service levels compared to combustion engines.

Thus, it is crucial for a redeployment policy to take battery levels and charging into account. MED is based on the concept of matching the anticipated energy demand at different stations with the expected energy supply at those stations. While the energy demand depends on the incidents and, consequently, the amount of energy needed to handle all incidents. On the other hand, the expected energy supply depends mainly on the distribution of ambulances across the base stations, which is influenced by redeployment decisions. Whenever a redeployment decision needs to be made, our approach deploys the ambulance to the station, which minimizes the energy deficit.

Our proposed method consists of three steps described in the remainder of this section:

(1) Determine the expected energy demand.
(2) Determine the expected energy supply.
(3) Calculate and minimize the energy deficit.

4.1 Expected Energy Demand

We introduce the concept of energy demand θ_w, which refers to the expected energy required to handle incoming incidents within the lookahead duration Δt at base station w. It is determined based on the expected number of incidents in the vicinity of the base station $d_w(t_{\text{now}}, \Delta t)$ during the lookahead duration and an expected energy use per incident ρ_w. The expected energy demand can be expressed as the product of these:

$$\theta_w = d_w(t_{\text{now}}, \Delta t)\rho_w \tag{1}$$

We define the demand forecast $d_w(t_{\text{now}}, \Delta t)$ as a function that estimates the expected number of incidents in the vicinity of the station w from the current time t_{now} until $t_{\text{now}} + \Delta t$. Numerous approaches have been proposed in the literature for predicting ambulance demand [21, 22, 25, 26]. These methods include but are not limited to machine learning techniques, time series analysis, and statistical models. In this paper, we compute an hourly historical average for demand prediction. [6] shows that this method yields a strong baseline for predicting ambulance demand. Let us note that our approach does not depend on a specific forecasting method and likely benefits from more accurate predictions. We leave the exploration of more sophisticated demand models to future research.

The vicinity V_w of a base station w is defined to be the incident locations $i \in V$ where the travel time $\tau(w, i)$ is shorter than from any other station. Mathematically, this can be expressed as follows:

$$V_w = \{i \in V | \tau(w, i) \leq \min_{w' \in W} \tau(w', i)\} \tag{2}$$

Using historical incident data, we calculate the average number of incidents per hour $\kappa_w(h)$ in the vicinity of each base station w and each hour of day $h \in \{0, ..., 23\}$. Let $\beta_h \in [0, 1]$ represent the fraction of hour h in the time interval $[t_{\text{now}}, t_{\text{now}} + \Delta t]$. The demand forecast is then given by:

$$d_w(t_{\text{now}}, \Delta t) = \sum_{h \in \{0, \dots, 23\}} \beta_h \kappa_w(h) \tag{3}$$

Determining the expected energy per incident within the proximity of each station holds significant importance. This necessitates evaluating the energy expenditure for traveling from a base station to the incident location, potentially to a hospital and returning to a station. A simplistic approach would assume that incidents solely occur at the centers of each demand area (i.e., the base stations), then travel to the nearest hospital, and finally return to the closest station. However, such an approach lacks accuracy. Therefore, we assume that the locations of incidents are uniformly spatially distributed across all possible incident locations $i \in V_w$ in the vicinity of w. We consider the probability of requiring transportation to a hospital, as well as accounting for the distribution of patients transported to different hospitals and the expected energy for redeployment to a station. The hospital distribution and the probability of requiring hospital transportation are derived from historical data.

We denote the proportion of incidents requiring hospital transportation as α, while α_h is the fraction of these incidents handled by hospital h. To calculate the expected energy use per incident ρ_w in the vicinity of station w, we first determine the expected driving time for fully handling an incident and redeployment to a station. Subsequently, we estimate the energy usage by multiplying the resulting driving times with the parameter $P_{driving}$, which approximates the energy consumed per unit of time:

$$\mathbb{E}(\rho_{hospital}(i)) = \sum_{h \in H} \alpha_h \frac{1}{|W|} \sum_{w' \in W} (\hat{\tau}(i, h) + \tau(h, w')) \tag{4a}$$

$$\mathbb{E}(\rho_{base}(i)) = \frac{1}{|W|} \sum_{w' \in W} \tau(i, w') \tag{4b}$$

$$\rho_w = P_{driving} \frac{1}{|V_w|} \sum_{i \in V_w} \hat{\tau}(w, i) + \alpha \mathbb{E}(\rho_{hospital}(i)) + (1-\alpha) \mathbb{E}(\rho_{base}(i)) \tag{4c}$$

4.2 Expected Energy Supply

This section focuses on outlining the methodology for calculating the expected energy supply ϕ_w, at a base station w over a specific time period. The actual energy supply depends on the demand, as ambulances may leave the base station to respond to incidents. While it is theoretically possible to model the distribution of incidents and to sample from an exponentially expanding set of future scenarios to derive estimates, finding optimal solutions is computationally intractable. Even approximations similar to the hindsight planning approaches in [24] are impracticable due to the inherent complexity and real-time constraints of the DEAR problem. To address this, we propose calculating an optimistic, expected energy supply $\hat{\phi}_w$, assuming that no incidents occur and no ambulances are redeployed during the prediction horizon, effectively disregarding the demand. This simplification allows for a deterministic calculation. However, we account for the probability of ambulances being dispatched and subsequently reducing the energy supply during the lookahead duration Δt. This is achieved by introducing a charging discount factor $\gamma \in [0, 1]$ to adjust the expected energy supply, resulting in $\phi_w = \gamma \hat{\phi}_w$. Note that even with those assumptions, determining the expected energy supply still requires simulating the complex charging logic and considering the arrivals of ambulances en route to the base station.

4.3 Minimize Energy Deficit

After we have defined the expected energy demand and supply, we continue by specifying how to calculate the energy deficits and subsequently dynamically redeploy ambulances. To define the energy deficit δ_w at a specific base station w, we calculate the difference between the expected energy demand and supply: $\delta_w = \theta_w - \phi_w$. However, simply minimizing this deficit has certain limitations. For instance, if a station already has sufficient supply to meet the demand, adding more supply would be unnecessary, even if it reduces the deficit. Therefore, we introduce a weighted deficit ω_w using a soft plus function [9]. This function assigns lower importance to stations with negative deficits (i.e., surplus supply compared to demand) and prioritizes stations with high deficits. The weighted deficit is calculated as follows:

$$\omega_w = \log(1 + exp(\frac{1}{100} \delta_w)) \tag{5}$$

In the last step, we describe the methodology for using the weighted energy deficit ω_w to make redeployment decisions. Whenever an ambulance a needs to be redeployed, we simulate sending the ambulance to each base station w to obtain $\omega_w(a)$. This is used to calculate the reduction in the expected weighted energy deficit $\omega_w - \omega_w(a)$ at each station. Subsequently, we redeploy the ambulance to the station that yields the most significant reduction.

4.4 Computational Complexity

Making ambulance redeployment decisions is a time-critical task, and any method should be able to compute a redeployment decision within seconds. Consequently, we designed our approach with this requirement in mind. To make each redeployment decision, we must assess the expected energy demand ρ_w and the expected energy supply ϕ_w at each station w both with and without the ambulance being redeployed. The energy demand consists of two components, the expected number of incidents and energy use per incident. The complexity of the former depends on the demand prediction model used. In this paper, we use the historical average, which can be pre-computed so that a prediction can be made in constant time. The second component, the expected energy use per incident, is a constant factor that can also be pre-computed. Therefore, calculating the expected energy demand has constant complexity. The primary computational effort lies in determining future energy supplies, which involves simulating the charging logic of each ambulance. As we assume optimistically that ambulances will not be deployed, they will eventually reach full charge, and their energy supply will no longer change. In other words, each ambulance adds a particular constant computational effort to simulate. From a computational point of view, the time complexity of determining future energy supplies is linear in the number of ambulances, resulting in a complexity of $O(|A|)$.

Our approach has an overall worst-case time complexity of $O(|A||W|)$. For each station, we need to calculate the expected energy demand (with constant complexity due to pre-computation) and compute the expected energy supply twice.

We implemented our method in C++ to obtain evaluation results presented in the next section. Executed on a notebook with Intel® Core™ i7-10750H CPU, one redeployment decision is obtained in approximately 0.23 milliseconds during a typical evaluation run with 45 base stations and 25 ambulances. Repeating the measurements with 1,000 ambulances in the environment (an unreasonably high number for benchmark purposes only), one decision is obtained in approx. 0.26 ms. These results satisfy the real-time requirement.

5 EVALUATION

In this section, we evaluate various solutions in a DEAR setting based on real-world emergency data from the city of San Francisco. We will first detail our experimental setup and, afterward, examine the impact of electric ambulances on DAR solutions and the performance of our newly proposed method *MED*.

5.1 Simulation environment

We evaluate various scenarios using an event-based simulator that replays real-world emergency data. This simulator mirrors the DEAR problem defined in Section 3 to simulate the operations of the EMS with electric ambulances. The foundation of our simulator is an openly accessible simulation environment for dynamic ambulance redeployment developed by [23]. Since this simulation does not consider electric vehicles, we extended it to include vehicles' battery state, charging, and energy use. Further, base stations were modified to contain a definable number of chargers of specified charging power with the problem definition's charging logic. Note that charging electric vehicles is a complex process influenced by factors such as battery level, battery condition, and ambient temperature. Similarly, energy usage depends on variables like driving profile, traffic conditions, and secondary loads such as heating or equipment required for patient care. Given the complexity of modeling these factors accurately, we simplify our simulation by utilizing constant values for charging power and driving energy usage, respectively.

The simulated EMS system is based on the city of San Francisco, USA. The system contains eleven hospitals and 45 base stations. Their locations are depicted in Figure 2. The road network graph used in the simulation was acquired from OpenStreetMap [1], with intersections representing the graph nodes. Hospitals and base stations were attached to the nearest node in the graph. Driving times

[1]ODbL license https://www.openstreetmap.org/copyright. Map data copyrighted by OpenStreetMap contributors and available from https://www.openstreetmap.org.

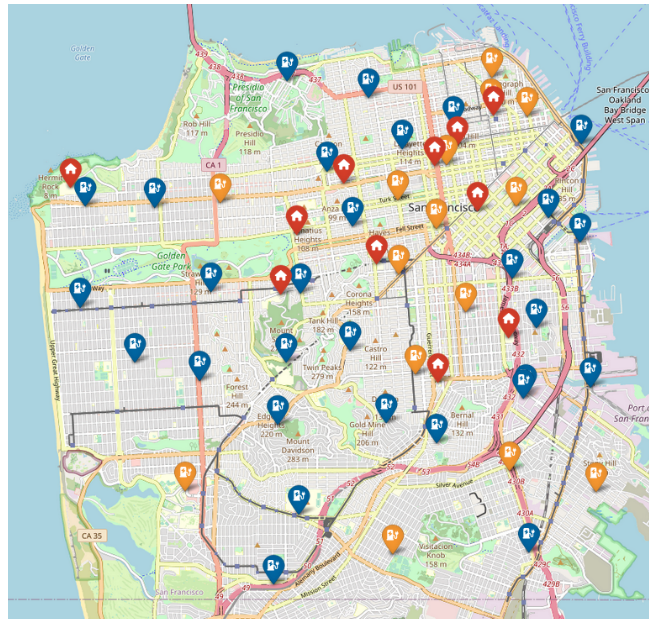


Figure 2: Simulation environment of San Francisco, USA. Locations of base stations are marked in orange if a fast charger is present and otherwise, in blue; Hospitals are marked in red. Note that population density is highest in the downtown area (top right). Map data © OpenStreetMap contributors.

were computed based on the shortest path with respective street limits depending on the road type. To account for traffic and slowing down due to turns and crossings, we calibrated the driving times based on estimates by HERE Traffic [2] by multiplying a constant factor. Based on this method, the average speed, including traffic congestion, was estimated to be 32 $\frac{km}{h}$. Ambulances returning to a base station are assumed to drive at traffic speed. However, when moving toward an incident or hospital, ambulances are granted certain exemptions from traffic regulations, allowing them to drive faster. Nevertheless, traffic congestion and safety considerations still limit realistic driving speed. Thus, we scaled driving times accordingly, resulting in an average emergency speed of 50 $\frac{km}{h}$ as suggested by [11].

The city of San Francisco has made real incident data publicly available through their *Fire Department Calls for Service* dataset [3]. This dataset contains historical records of health emergency calls, including information such as the date, time, and location of each emergency. This enables us to simulate the historical occurrence of incidents with arbitrary configurations of base stations, ambulances, and redeployment methods. However, it is important to note that while the dataset indicates whether a hospital was targeted, specific details such as the hospital's name or location are not disclosed. Selecting a suitable hospital involves a complex decision process, including various factors such as the patient's medical needs, hospital occupancy levels, patient preferences, and the proximity to hospitals [1]. Since this information is unavailable to us,

[2]https://www.here.com/platform/traffic-solutions/real-time-traffic-information
[3]https://data.sfgov.org/Public-Safety/Fire-Department-Calls-for-Service/nuek-vuh3

we determine target hospitals by random sampling according to the real-world distribution of patient transports to hospitals between February 2022 and February 2023 published by the Data Working Group (DWG) of the City of San Francisco [4]. Note that this random sampling was done as a preprocessing step, ensuring the hospital transportations are consistent across all experiments. Locations of incidents were mapped to the nearest graph node in our simulation.

5.1.1 Placement and power of fast chargers. As discussed in our introduction, cities will likely outfit only a subset of base stations with fast chargers, primarily due to installation costs. Therefore, we strategically locate fast chargers at the stations with the highest demand, assigning one fast charger per station. Note that according to our problem definition, additional slower chargers are already present at every station. In our evaluation, we considered three different types of fast chargers, each offering different charging powers. The first type is a high-power DC charger delivering 50 kW of charging power. The second type is a cheaper three-phase AC charger delivering 22 kW charging power. Lastly, we considered a more expensive option of 100 kW charging.

5.1.2 Electric Ambulance Models. The battery capacity and average driving energy use (P_{driving}) in our simulation are based on real-world electric ambulances. We base our experiment values on electric ambulance "WAS 500" because it is a suitable replacement for ICE ambulances and technical data is readily available [5]. We set the battery capacity to 87 kWh, based on the specifications provided in the datasheet of the ambulance. To determine P_{driving}, we consider the average speed and the energy usage from the datasheet. We calculate this value as 30 kW.

5.2 Metrics

As motivated in our introduction, minimizing ambulance response times is critical for EMS providers. In an ambulance redeployment context, response times are usually defined as the time between dispatching an ambulance at its base station and its arrival at the incident scene. Aggregated metrics used for evaluating the performance of EMS systems are the average response time (**ART**) and the fraction of response times within a certain response time threshold (RTT) [17, 23]. RTT values and targeted fractions are set differently by different institutions [17]. San Francisco's Emergency Medical Services Agency aims to arrive at life-threatening incidents within a 10 minute threshold at least 90% of the time [2, 23]. We use this metric extensively in our evaluation, denoting it as **RTT10**. We occasionally also include RTT fractions for 8 minutes (**RTT8**) and 12 minutes (**RTT12**).

5.3 Baselines

We compare our method **MED** (Minimize Energy Deficit) with several straightforward baselines as well as several state-of-the-art approaches for redeploying combustion engine ambulances. The most simple baseline is **RAND**, which redeploys the ambulance to a random base station. **NEAR** selects the base station which can be reached fastest by the ambulance (i.e. minimizes driving time). **NEARC** and **NEARF** similarly select the nearest station but consider only stations with chargers (NEARC) or free, fast chargers (NEARF), respectively. Note that this method checks availability at query time. We also include state-of-the-art approaches from the DAR problem discussed in (Section 2) and refer to them as conventional approaches. These approaches consists of static methods, namely **ERTM**[3] and **MEXCLP**[8, 13], a dynamic method called **DMEXCLP**[13], and the reinforcement learning based approach **DRLSN**[14]. Let us note that *DRLSN* is trained in an environment considering DEAR, and thus, it can learn the specific behavior of E-Ambulances. However, we did not change the agent itself as a straightforward extension of observation data did not yield improved results.

[4] http://sfemergencymedicalresponse.weebly.com/ambulance-destinations.html
[5] https://www.was-vehicles.com/en/innovation/was-500-electric-ambulance.html

Table 1: RTT10 performance of conventional methods in the ICE case compared to the EV case with different charging powers and 24 ambulances.

Scenario	ERTM	DRLSN	MEXCLP	DMEXCLP
ICE	0.88	**0.90**	0.83	0.89
EV 22 kW	**0.47**	0.40	0.30	0.20
EV 50 kW	0.72	**0.85**	0.58	0.57
EV 100 kW	0.76	**0.86**	0.67	0.56

5.4 Results

In this section, we present the results of our experimental evaluation based on the previously described simulation environment to answer the following research questions:

(1) How large is the effect of replacing ICE ambulances with EVs using established DAR methods?
(2) Does our approach *MED* perform better than methods from related work for DEAR?
(3) What is the influence of simulation parameters such as the number of available chargers?
(4) How sensitive is our approach to variation of its parameters?

For all experiments, methods were evaluated by simulating one year of incidents (*test set*) in our simulation. The resulting response times were then aggregated to obtain RTT10 and ART metrics. The respective previous year (*validation set*) was used to determine the method's parameters, such as historical demand and selecting hyper-parameters. The best set of hyper-parameters (according to the RTT10 metric) was selected for evaluation on the test set. Experiments were conducted for the years 2015 to 2022. Due to the numerous parameters involved, including different combinations of years, ambulance quantities, charger quantities, charging power, etc., we cannot present all results here. Unless indicated otherwise, the experiments were conducted with incidents from the year 2022, using 15 fast chargers, each providing 50 kW charging power. Additionally, we included variations of these parameters to facilitate a comprehensive comparison of methods under different scenarios.

5.4.1 Effect of switching to electric ambulances. In this section, we analyze the effectiveness of methods for ordinary DAR settings (conventional approaches) when being applied to the DEAR problem. We present the results for ICE and EV scenarios containing 24 ambulances in Table 1, as 24 ambulances are required for the

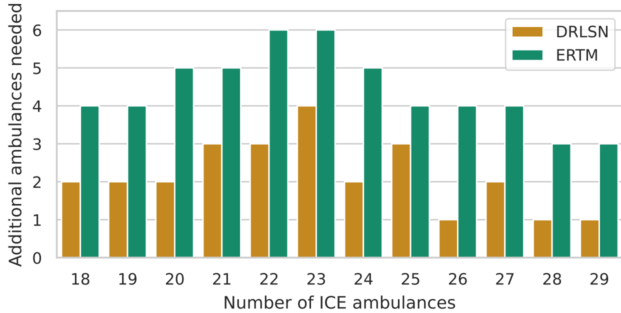


Figure 3: Number of additional ambulances needed to reach same performance (RTT10 metric) as in the non-EV scenario for best methods from related work. 50 kW charging power.

first method to reach the RTT10 target of 90%. We observe a significant decline in the RTT10 metric when introducing energy use and charging, with some cases showing a reduction of more than 50% in performance. Using 22 kW fast chargers results in inferior performance: *ERTM* receives the best 22 kW RTT10 score (0.47), which is not acceptable for an EMS provider, despite its significantly better performance (0.88) in the ICE case. When using 50 kW or 100 kW fast chargers, the decline in performance is less severe but still substantial. *DRLSN* achieves best RTT10 scores in the ICE (0.90), 100 kW (0.86) and 50 kW (0.85) cases. Notably, its reward-based algorithm shows the ability to learn certain characteristics of the EV environment despite not explicitly observing energy-related data. Its poor performance in the 22 kW case may be explained by rewards being too sparse to enable effective training. Like *ERTM*, both *MEXCLP* and *DMEXCLP* show drastic decreases in performance. Although the dynamic method *DMEXCLP* performs better than the static approaches *ERTM* and *MEXCLP* in the ICE case, it experiences substantial difficulties in the EV scenarios, even showing worse results in the 100 kW case compared to 50 kW. Overall, results indicate that using fast chargers with 22 kW charging power will not enable acceptable performance with these methods. Increasing the charging power to 50 kW improved the results, but additional ambulances are still necessary. Installing 100 kW chargers does not appear to improve results substantially. As infrastructure investments generally increase with higher charging power, emergency medical service providers should be aware of this effect when transitioning to electric ambulances.

Figure 3 provides insights into the number of additional ambulances needed when transitioning from ICE to EV ambulances. It depicts the number of additional ambulances required to reach an equal or better RTT10 performance compared to non-EV ambulances for the *ERTM* and *DRLSN*. We use 50 kW chargers in the scenario, as there is a minimal improvement when using 100 kW. *ERTM* requires an additional 3 to 6 ambulances. *DRLSN* requires 2 to 4 additional ambulances when replacing up to 25 ICE ambulances. In settings replacing more than 25 ICE the number of additional ambulances can decrease to 1.

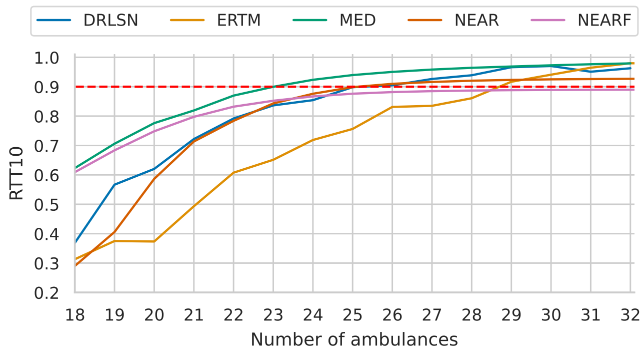
Overall, our results show that employing conventional methods from related work on the DEAR problem requires more ambulances to achieve a similar level of performance compared to ICE ambulances. Additionally, an interesting finding is that the difference between 50 kW and 100 kW charging is minimal in contrast to charging with 22 kW.

Table 2: Performance of all methods when using 24 ambulances and 50 kW charging power.

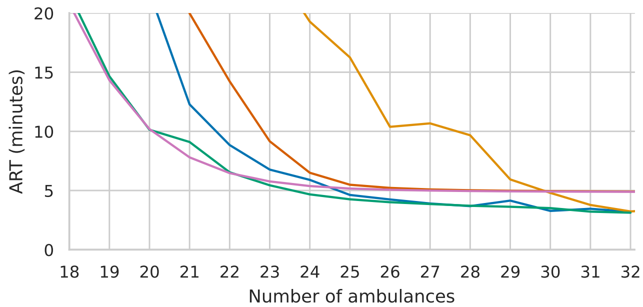
Method	RTT8	RTT10	ART
MED	**0.87**	**0.92**	**4.64**
NEAR	0.79	0.88	6.50
NEARF	0.79	0.87	5.38
DRLSN	0.81	0.85	5.90
NEARC	0.75	0.84	5.89
ERTM	0.68	0.72	19.27
MEXCLP	0.50	0.58	32.44
DMEXCLP	0.49	0.57	33.87
RAND	0.01	0.01	153.09

5.4.2 Performance of MED. We now introduce results for our approach *MED* and compare them to state-of-the-art conventional methods developed for DAR, as well as our DEAR baselines. Results for all methods are shown in Table 2. We again chose 24 ambulances and 50kW charging power due to the previously mentioned practical relevance of this scenario. Our approach *MED* outperforms all other methods across all metrics. Specifically, it achieves an RTT10 value of 0.92, which is well within the 90% target. The average response time (ART) of 4.64 minutes is about 45s faster than the second-best method *NEARF*, and 75s less than *DRLSN*, the best conventional method from related work. It is worth noting that another nearest station method, *NEAR*, also demonstrates surprisingly good performance, securing the second-best RTT10 value of 0.88. The best performing conventional method is *DRLSN* (0.85), followed by *ERTM* (0.72). The difference between RTT10 and ART scores, especially when considering the comparatively good performance of simplistic baselines such as *NEAR* or *NEARF* underlines the observation that conventional methods do not perform well in the evaluated EV scenario. In contrast, *MED* performs better in DEAR (RTT10 of 0.92) than the best DAR approach in the corresponding ICE scenario (RTT10 of 0.90, compare Figure 1).

The relationship between the number of deployed ambulances and the performance is illustrated in Figure 4 for the best-performing methods. *MED* consistently demonstrates strong results across all metrics. Analyzing the RTT10 performance in Figure 4a, it becomes evident that *MED* outperforms other approaches with a substantial gap to the second best method up to a number of 29 ambulances. As noted before, it is the first to exceed the 90% RTT10 target (dashed red line). Furthermore, its performance considering the ART metric (Figure 4b) is superior to others in the most interesting region (due to its closeness to the 90% RTT10 target) of about 24 ambulances. When 22 or fewer ambulances are used, method *NEARF* yields lower ART values. This is because in these cases, demand for ambulances, and the energy use that comes with it, is so high that all other objectives fade in comparison to obtaining energy as fast as possible. As the *NEARF* method is designed to immediately drive to the nearest free charger, regardless of its location or any



(a) RTT10 metric. Dashed line indicates 90% target.



(b) ART metric.

Figure 4: Performance comparison of best methods for 50 kW charging power.

Table 3: Number of ambulances needed to reach the 90% RTT target for various RTT values. 50 kW charging power.

Method	8 min	10 min	12 min
MED	**26**	**24**	**22**
DRLSN	29	26	25
NEAR	> 40	26	24
ERTM	30	29	29
DMEXCLP	> 40	32	31
MEXCLP	> 40	32	29
NEARC	> 40	> 40	23
NEARF	> 40	> 40	23
RAND	> 40	> 40	> 40

Table 4: Performances of MED compared to best method from related work for each evaluation year. In each year, MED performed best, followed by DRLSN. The number of ambulances in each row was determined as the lowest amount that reached 90% RTT10 for the given year. Column *Diff* for RTT10 is the decrease of incidents that could not be reached within 10 minutes. Column *Diff* for ART is the decrease in response times.

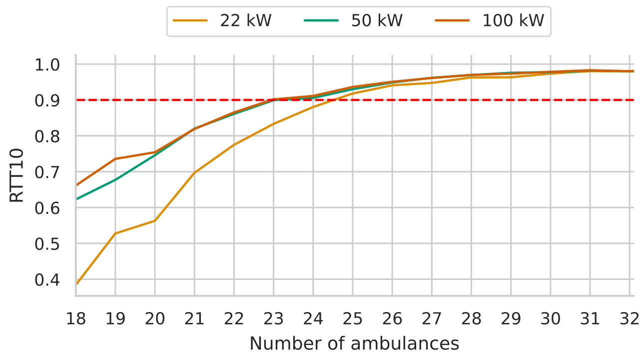
Year	RTT10			ART		
	MED	DRLSN	Diff	**MED**	DRLSN	Diff
2015	0.901	0.860	-29.18%	4.810	5.551	-13.34%
2016	0.907	0.867	-30.57%	5.101	5.916	-13.78%
2017	0.919	0.877	-34.28%	4.494	5.345	-15.92%
2018	0.912	0.867	-33.63%	4.679	5.491	-14.79%
2019	0.910	0.875	-28.19%	4.579	5.469	-16.27%
2020	0.908	0.872	-28.18%	4.635	5.476	-15.37%
2021	0.905	0.852	-35.42%	4.927	5.767	-14.57%

other criteria, it fulfills this objective well. In situations where a substantial number of ambulances (27 or more) are available, methods from previous research narrow the gap. At this point, the locations and availability of chargers become less critical as it becomes more likely that a charged ambulance is stationed sufficiently close to any incident. Furthermore, slow charging is sufficient to make sure that drained ambulances will be available at a later point in time. It is, however, interesting that the gap for the RTT10 metric (c.f. Figure 4a) closes more slowly that the gap in ART (c.f. Figure 4b). This indicates that *MED* still allows significantly fewer incidents that are not handled within the 10-minute limit than compared methods up to 29 ambulances.

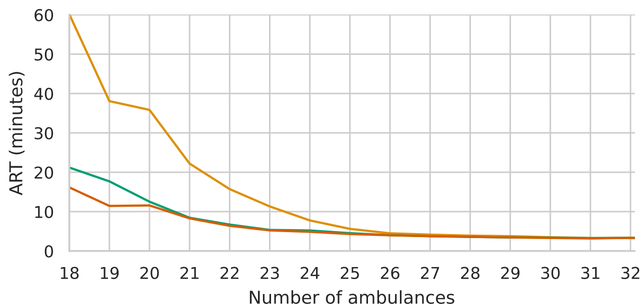
As emergency service providers usually aim to fulfill a certain minimum service level, we provide the number of ambulances needed to reach a 90% fraction of common RTT values in Table 3. An important observation is that *MED* requires the lowest number of ambulances to reach the target in all cases. A RTT8 target is reached by deploying 26 ambulances with *MED*, whereas the second best method, *DRLSN*, requires 29 ambulances. The RTT10 and RTT12 targets are reached with 24 and 22 ambulances, respectively, requiring two and one ambulances less than the runner-up. It is worth mentioning that most methods failed to reach the RTT8 target for fleet sizes up to 40, which is the maximum number of ambulances considered in our experiments.

To see if the superior performance of *MED* can be reproduced in other years, we repeated the experiment above for each pair of years starting in 2015. This includes fitting parameters on the given year and testing methods' performance in the following year. The results summarized in Table 4 show that *MED* can reach the 90% RTT10 target with fewer ambulances than methods from related work each year. The difference in incidents that could not be reached within 10 minutes is considerably lower in these cases, namely between 28.18% to 35.42% lower. Average response times also decrease consistently for all years. In absolute numbers, this means reducing average response times by about 50 seconds in our experiments, which can be valuable in critical emergencies.

These results demonstrate the superior performance of *MED* for the DEAR problem across various scenarios. Furthermore, as *MED* in DEAR displays a similar or better performance than compared methods in the ordinary DAR environments based on ICE ambulances, we can conclude that switching to an equally-sized fleet of E-Ambulances can be done without significantly decreasing response times.



(a) RTT10 metric.



(b) ART metric.

Figure 5: Performance of MED for different charging powers.

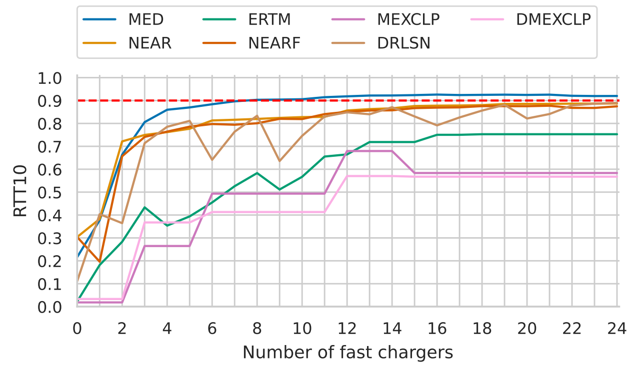


Figure 6: Comparison of RTT10 performance for different numbers of fast chargers. Scenario with 24 ambulances and 50 kW charging power.

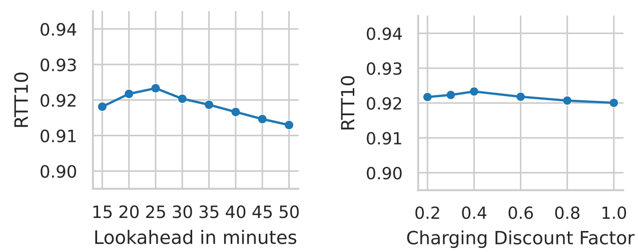


Figure 7: Performance of MED for varying hyper-parameters. Scenario with 24 ambulances and 50 kW charging power.

5.4.3 Varying power and number of chargers. The performance of *MED* for different charging rates and numbers of deployed ambulances is presented in Figure 5. It can be seen again that the difference between 50 kW and 100 kW fast charging power is minimal. However, 22 kW charging results in inferior performance. For example, the RTT10 target is reached with 25 ambulances instead of only 24 for both higher charging rates. This disparity becomes more apparent as the number of ambulances decreases, as the per-ambulance energy use and corresponding charging activity increases in such scenarios. With increasing numbers of ambulances, the charging pressure vanishes, which can be seen in the convergence of all powers' measurements. Figure 6 depicts the results of different methods for varying the number of installed fast chargers. As before, we use 24 ambulances as the lowest amount to be sufficient to reach the 90% RTT10 target. *ERTM*, *MEXCLP*, and *DMEXCLP* exhibit a slow increase of performance when increasing the number of fast chargers and thus appear to be especially ill-suited for the EV scenario. In contrast, the performance of *MED*, *NEAR*, *NEARF* and *DRLSN* follows an early quick increase with a slower rise once about three fast chargers are installed, i.e., they appear to either use less energy or utilize fast chargers better, or both. It should be noted that *MED* is the only approach that meets the 90% RTT10 level. Furthermore, *MED*'s performance does not substantially increase when more than 11 chargers are installed in the environment. To summarize, our method tailored for EV scenarios requires not only fewer ambulances but also fewer fast chargers.

5.4.4 Parameter sensitivity. Figure 7 shows how varying *MED*'s two parameters affect its RTT10 performance, using the scenario of 24 ambulances and 50 kW charging. Examining the parameter lookahead duration Δt (Figure 7 left), the optimal value is 25 minutes, with a roughly linear decrease when higher or lower values are used. The sensitivity of our approach to this parameter is low, as doubling it to 50 minutes only marginally decreases RTT10 performance by about 0.01. Varying the charging discount factor γ (Figure 7 right) appears to have little effect on performance. The optimum is at a value of 0.4, which can be explained by charging processes at base stations being frequently interrupted due to incoming incidents in this challenging scenario.

5.4.5 Qualitative analysis. Figure 8 shows a snapshot of our simulation from the point of view of our approach *MED*. The weights assigned by the method (orange bars) are calculated in a way that expected demand (red bars) is offset by available energy (blue bars), i.e., ambulances assigned to the respective base station. The energy distribution appears to be pretty spread out to minimize response times. Several base stations necessarily contain zero energy because, in this scenario, 25 ambulances have to cover all 45 base stations. However, the gaps are mostly in lower demand areas and can be covered by nearby base stations with assigned ambulances.

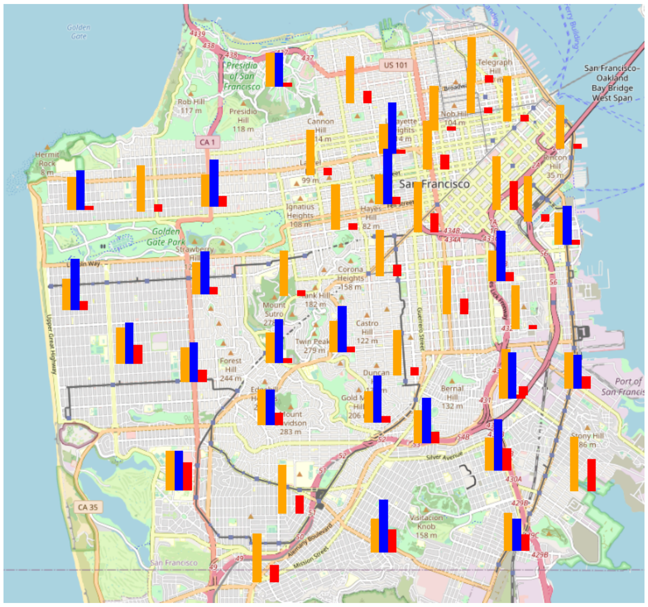


Figure 8: Snapshot of a simulation with 25 ambulances. Bar plots indicate base stations' estimated future energy values as calculated by MED: Supply (blue); Demand (red); Deficits (yellow) after nonlinear scaling (higher scores mean higher priority). Map data © OpenStreetMap contributors.

6 CONCLUSION

In this paper, we introduce the Dynamic Electric Ambulance Redeployment (DEAR) problem extending the Dynamic Ambulance Redeployment (DAR) problem to electric ambulances. We propose the Minimize Energy Deficits (MED) method, which determines redeployment actions by estimating the future energy deficit over all base stations. The energy deficit of a base station weighs a prediction of future demand against a prediction of the available energy level corresponding to the remaining range of stationed ambulances. We conducted experiments in a realistic scenario using an event-based simulator based on real-world incidents. Results show that MED reaches better performance than compared DAR methods, as well as baselines for EV settings. Furthermore, our results indicate that transitioning to electric ambulances can be done without increasing the number of available ambulances while maintaining comparable response times.

For future work, we plan to explore using more sophisticated prediction methods for demand and available energy. Furthermore, we want to examine sequential planning approaches considering multiple decisions in advance.

REFERENCES

[1] San Francisco Emergency Medical Services Agency. [n. d.]. *San Francisco EMS Ambulance Destinations.* http://sfemergencymedicalresponse.weebly.com/ambulance-destinations.html

[2] San Francisco Emergency Medical Services Agency. 2011. *Prehospital Provider Standards (Policy 4000).* https://www.sfdph.org/dph/files/EMS/Policy-Protocol-Manuals/Policy-Manual/1538-4000ProviderStandards09-01-2011.pdf

[3] Pieter L Van Den Berg and J Theresia Van Essen. 2019. Comparison of static ambulance location models. *International Journal of Logistics Systems and Management* 32, 3-4 (2019), 292–321.

[4] Lawrence H Brown, Christa L Whitney, Richard C Hunt, Michael Addario, and Troy Hogue. 2000. Do warning lights and sirens reduce ambulance response times? *Prehospital Emergency Care* 4, 1 (2000), 70–74.

[5] Andreas Bürger, Jan Wnent, Andreas Bohn, Tanja Jantzen, Sigrid Brenner, Rolf Lefering, Stephan Seewald, Jan-Thorsten Gräsner, and Matthias Fischer. 2018. The effect of ambulance response time on survival following out-of-hospital cardiac arrest: an analysis from the German resuscitation registry. *Deutsches Ärzteblatt International* 115, 33-34 (2018), 541.

[6] Nabil Channouf, Pierre L'Ecuyer, Armann Ingolfsson, and Athanassios N Avramidis. 2007. The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta. *Health care management science* 10 (2007), 25–45.

[7] Richard Church and Charles ReVelle. 1974. The maximal covering location problem. In *Papers of the regional science association*, Vol. 32. Springer-Verlag Berlin/Heidelberg, 101–118.

[8] Mark S Daskin. 1983. A maximum expected covering location model: formulation, properties and heuristic solution. *Transportation science* 17, 1 (1983), 48–70.

[9] C Dugas, Y Bengio, F Bélisle, and C Nadeau. 2001. Incorporating second order functional knowledge into learning algorithms. *Advances in Neural Information Processing Systems* 13 (2001), 472–478.

[10] Shakiba Enayati, Maria E Mayorga, Hari K Rajagopalan, and Cem Saydam. 2018. Real-time ambulance redeployment approach to improve service coverage with fair and restricted workload for EMS providers. *Omega* 79 (2018), 67–80.

[11] Michel Gendreau, Gilbert Laporte, and Frédéric Semet. 2001. A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. *Parallel computing* 27, 12 (2001), 1641–1653.

[12] Jeffrey Goldberg, Robert Dietrich, Jen Ming Chen, M George Mitwasi, Terry Valenzuela, and Elizabeth Criss. 1990. Validating and applying a model for locating emergency medical vehicles in Tuczon, AZ. *European Journal of Operational Research* 49, 3 (1990), 308–324.

[13] Caroline J Jagtenberg, Sandjai Bhulai, and Robert D van der Mei. 2015. An efficient heuristic for real-time ambulance redeployment. *Operations Research for Health Care* 4 (2015), 27–35.

[14] Shenggong Ji, Yu Zheng, Zhaoyuan Wang, and Tianrui Li. 2019. A deep reinforcement learning-enabled dynamic redeployment system for mobile ambulances. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–20.

[15] Olive C Kobusingye, Adnan A Hyder, David Bishai, Manjul Joshipura, Eduardo Romero Hicks, and Charles Mock. 2006. Emergency medical services. *Disease Control Priorities in Developing Countries. 2nd edition* (2006).

[16] Matthew S Maxwell, Mateo Restrepo, Shane G Henderson, and Huseyin Topaloglu. 2010. Approximate dynamic programming for ambulance redeployment. *INFORMS Journal on Computing* 22, 2 (2010), 266–281.

[17] Laura A McLay and Maria E Mayorga. 2010. Evaluating emergency medical service performance measures. *Health care management science* 13 (2010), 124–136.

[18] Jill P Pell, Jane M Sirel, Andrew K Marsden, Ian Ford, and Stuart M Cobbe. 2001. Effect of reducing ambulance response times on deaths from out of hospital cardiac arrest: cohort study. *Bmj* 322, 7299 (2001), 1385–1388.

[19] Charles ReVelle and Kathleen Hogan. 1989. The maximum availability location problem. *Transportation science* 23, 3 (1989), 192–200.

[20] Emmanouil S Rigas, Antonis Billis, and Panagiotis D Bamidis. 2022. Can Artificial Intelligence Enable the Transition to Electric Ambulances? In *Challenges of Trustable AI and Added-Value on Health.* IOS Press, 73–77.

[21] Hubert Setzler, Cem Saydam, and Sungjune Park. 2009. EMS call volume predictions: A comparative study. *Computers & Operations Research* 36, 6 (2009), 1843–1851.

[22] Krisjanis Steins, Niki Matinrad, and Tobias Granberg. 2019. Forecasting the demand for emergency medical services. (2019).

[23] Niklas Strauß, Max Berrendorf, Tom Haider, and Matthias Schubert. 2022. A Comparison of Ambulance Redeployment Systems on Real-World Data. In *Proceedings of the 1st Workshop on Urban Internet-of-Things Intelligence (UNIT 2022) co-located with the 22nd IEEE International Conference on Data Mining (ICDM 2022).*

[24] Niklas Strauß, Lukas Rottkamp, Sebastian Schmoll, and Matthias Schubert. 2021. Efficient Parking Search using Shared Fleet Data. In *2021 22nd IEEE International Conference on Mobile Data Management (MDM).* IEEE, 115–120.

[25] Zhaonan Wang, Tianqi Xia, Renhe Jiang, Xin Liu, Kyoung-Sook Kim, Xuan Song, and Ryosuke Shibasaki. 2021. Forecasting ambulance demand with profiled human mobility via heterogeneous multi-graph neural networks. In *2021 IEEE 37th International Conference on Data Engineering (ICDE).* IEEE, 1751–1762.

[26] Zhengyi Zhou and David S Matteson. 2015. Predicting ambulance demand: A spatio-temporal kernel approach. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining.* 2297–2303.