Part II

Offline Approaches

5

The Minimum Expected Penalty Relocation Problem for Ambulance Compliance Tables

The previous chapters were concerned with the online approach to the ambulance relocation problem: the majority of the computational effort is done at the decision moment itself. From this chapter onwards, we shift our focus to the offline approach, in which most computations are done a priori and the solutions are stored. When a certain situation occurs, the solution is consulted and applied, possibly proceeded by an additional short computation. Whereas online policies can take into account many characteristics of the state of the EMS system, this is impractical for offline methods as this would induce many system states for which a solution have to be computed in advance. Therefore, offline methods use little information about the state of the system.

A commonly used offline strategy is the *compliance table* policy. The system state in a compliance table is purely given by the number of available vehicles: each compliance table level indicates the desired waiting site locations for the available ambulances. If these ambulances are at their desired waiting sites, the system is said to be *in compliance*. The number of available ambulances changes when a request arrives or when an ambulance becomes available again. Then, each idle ambulance may be assigned to a different waiting site. As the number of units is bounded by the fleet size, computation of efficient compliance tables can be done offline.

To this end, we introduce the minimum expected penalty relocation problem (MEXPREP) in this chapter. In this problem, which we formulate as an integer linear program, one has the ability to control the number of waiting site relocations. Moreover, different performance measures related to response times, such as survival probabilities, can be incorporated. We show by simulation that the MEXPREP compliance tables outperform both the static policy and compliance tables obtained by the maximum covering relocation problem (MECRP), which both serve as benchmarks. Besides, we perform a study on different relocation thresholds and on two different methods to assign available ambulances to desired waiting sites.

This chapter is based on Van Barneveld (2016).

5.1 Introduction

In addition to the ability to calculate compliance tables offline, another important strength of the compliance table policy is that it is simple to explain to and to use by dispatchers, since the state of the EMS system is only described by the number of available ambulances. However, surprisingly little has been published about ambulance compliance tables, despite some practical advantages over online methods, as mentioned above. To the best of our knowledge, Gendreau et al. (2006) were the first to propose a methodology for computing compliance tables, by formulating the MECRP. Although the MECRP, which we will in summarize in Section 5.2.1 below, is a good and easily applicable model to compute compliance tables, it has some major limitations:

1. An area is covered if an idle ambulance is present within the coverage radius: multiple idle ambulances within the coverage radius do not contribute to the coverage of the area. Especially in an EMS system with a high call arrival rate, it may happen that another incident occurs before the idle ambulances reach the locations to which they are assigned, according to the compliance table. The MECRP does not take this into account: it only focuses on the next future emergency request. In other words, the MECRP utilizes the notion of single coverage, whereas probabilistic coverage may benefit the compliance table.
2. There are at least as many waiting site locations as ambulances. This is a rather strong assumption and not generally true in practice, although it tends to be more and more common in the US and Canada to park up (temporarily) at a street corner or other strategic hotspot. However, it may be dictated by law that ambulances are allowed to idle at designated ambulance base stations only.
3. The capacity of each waiting site location equals one. This may be true for designated ambulance parking spaces, but in general not for base stations.
4. Only a performance measure related to coverage can be incorporated.

As a consequence of limitations 1 and 3, each waiting site location occurs at most once in each compliance table level. However, it could be beneficial to locate multiple ambulances at a waiting site, e.g., at a waiting site in the middle of a densely populated area with a high call arrival rate, in order to anticipate a possible rapid succession of incidents occurring in that area. In addition, we are forced to do this in a system in which limitation 2 does not hold. We extend the MECRP in such a way that within a compliance table level, a waiting site can occur multiple times. We do this by incorporating the objective function of the maximum expected coverage location problem (MEXCLP), presented by Daskin (1983), into the objective function of the MECRP.

The last limitation is related to coverage. As pointed out by De Maio et al. (2003), the most common EMS standard is to respond to 90% of all urgent calls within 8 minutes. Many EMS systems use the percentage of calls covered as performance measure. However, as stated by Erkut et al. (2008), the black-and-white nature of the coverage concept is an important limitation, and standard coverage models should not be used for ambulance location. First, coverage can result in large measurement errors because of their limited ability to discriminate between different response times. Second, these measurement errors are likely to result in large optimality errors when one uses covering models to locate ambulances instead of a model that takes survival probabilities into account. The difference between 'coverage' and 'survival' is demonstrated by an artifical example by Erkut et al. (2008), and it is shown that covering models can result in arbitrarily poor location decisions for ambulances.

In the MECRP only the performance measure of coverage can be incorporated. The MEXPREP we propose in this paper, is an extension of the MECRP in which a general performance measure can be incorporated, including the concept of survival mentioned above. We do this by introducing a penalty function, which is a non-decreasing function that solely depends on the response time (hence the name minimal expected penalty relocation problem).

The remainder of this paper is organized as follows. In Section 5.2.1 we explain the MECRP of Gendreau et al. (2006). In Sections 5.2.2 and 5.2.3 we treat the limitations mentioned above, resulting in the formulation of the MEXPREP in Section 5.2.4. In Section 5.3, we consider two models for the assignment problem, which needs to be solved to obtain an assignment of available ambulances to the waiting sites corresponding to the compliance table level. We conclude the paper by a numerical study in Section 5.4.

5.2 Model

One method to compute compliance tables is solving the MECRP, presented by Gendreau et al. (2006). In this section, we will extend MECRP. Next, we proceed with a summary of this problem.

5.2.1 Maximal Expected Coverage Relocation Problem

The MECRP is defined on a directed graph $G = (V \cup W, A)$ representing the region of interest. The region is discretized into demand zones, e.g., postal codes, in which V is the vertex set of these demand points. Moreover, W is the vertex set of potential waiting sites for n emergency vehicles and A is a set of arcs defined on $(V \cup W)^2$. A travel time is associated to each arc $(i, j) \in A$ and d_i denotes the demand at vertex $i \in V$. This d_i may, for instance, correspond to the population of demand zone i, or to the probability that an incoming emergency call occurs in demand zone i, which can be estimated by analyzing historical data. A vertex i is said to be covered by a vertex $j \in W$ if the expected travel time from j to i, denoted by τ_{ji}, is less than a given coverage radius T, expressed in time. We denote by W_i the subset of vertices of W covering i. We refer to Table 3.1 for an overview of the used notation.

In MECRP, the *busy fraction* p plays an important role. This is the probability that an ambulance is busy, i.e., responding to an emergency call or serving or transporting a patient. This busy fraction could be computed by $p = \lambda/(n\mu)$, where λ is the call arrival rate, μ is the average service rate and the number of ambulances is n. This busy fraction may also be estimated by analysis of historical data. The probability of being in a situation with k available ambulances, denoted by π_k, can easily be computed by, for instance, means of a binomial distribution:

$$\pi_k = \binom{n}{k}(1-p)^k p^{n-k}, \quad k = 0, 1, \ldots, n. \tag{5.1}$$

As was pointed out by Gendreau et al. (2006), a simple relaxation procedure for the MECRP consists of solving the MCLP, presented by Church and ReVelle (1974), for each compliance table level $k = 1, \ldots, n$. This procedure produces a compliance table, but it ignores constraints on waiting site changes at each event. To incorporate such constraints, it is useful to view the system as being in a succession of states k over time, where k is the number of available ambulances. In the remainder, we will call the row of the compliance table level with k waiting sites, the k^{th} level of the compliance table, which indicates the desired waiting sites for k available ambulances. This compliance table level k is described by binary variables x_{jk} equal to 1 if and only if an ambulance is located at $j \in W$, and by binary variables y_{ik} equal to 1 if demand point i is covered by at least one ambulance in compliance table level k. Moreover, a bound α_k is imposed on the number of waiting site changes between compliance table levels k and $k+1$, where $1 \leq k \leq n-1$. As a consequence, binary variables a_{jk} are defined, which equal 1 if and only if $j \in W$ ceases to be a waiting site in compliance table level $k+1$, starting from level k. The MECRP is formulated as follows:

$$\text{MECRP: Maximize} \sum_{k=1}^{n} \sum_{i \in V} d_i \pi_k y_{ik} \tag{5.2}$$

$$\text{Subject to:} \sum_{j \in W_i} x_{jk} \geq y_{ik} \qquad i \in V,\ k = 0, 1, \ldots, n \tag{5.3}$$

$$\sum_{j \in W} x_{jk} = k \qquad k = 0, 1, \ldots, n \tag{5.4}$$

$$x_{jk} - x_{j,k+1} \leq a_{jk} \qquad j \in W,\ k = 1, \ldots, n-1 \tag{5.5}$$

$$\sum_{j \in W} a_{jk} \leq \alpha_k \qquad k = 1, \ldots, n-1 \tag{5.6}$$

$$x_{jk} \in \{0, 1\} \qquad j \in W,\ k = 0, 1, \ldots, n \tag{5.7}$$

$$y_{ik} \in \{0, 1\} \qquad i \in V,\ k = 0, 1, \ldots, n \tag{5.8}$$

$$a_{jk} \in \{0, 1\} \qquad j \in W,\ k = 1, \ldots, n-1. \tag{5.9}$$

In this model, the objective function (5.2) maximizes the expected coverage. Constraints (5.3) induce that vertex $i \in V$ is covered only if at least one ambulance is located at at least one of the waiting sites in W_i, in compliance table level

k. Constraints (5.4) ensure that exactly k waiting sites are occupied in compliance table level k. Constraints (5.5) and (5.6) control the number of waiting site changes between compliance table levels k and $k+1$. The designated waiting sites at compliance table level k are given by decision variables x_{jk}. Although $k = 0$ is included in the original MECRP by Gendreau et al. (2006), it is not necessary to include this case.

5.2.2 Expected Covered Demand

In the MECRP, the objective function for a given compliance table level k is to maximize the demand covered within the response time threshold. Then, each level is weighted according to π_k, the probability of being in a situation with k available ambulances, which can be computed using Equation (5.1). As stated by Gendreau et al. (2006), the MECRP reduces to the MCLP with k ambulances if $\pi_k = 1$. After all, always k ambulances are available, since $\pi_i = 0$ for $i \neq k$.

Although the MCLP is a useful method for determining ambulance base locations, it has a major shortcoming: it assumes there is always an ambulance available at a base location. In practice, this is not true, since ambulances may be busy serving a patient. The fraction of duty time an ambulance is busy serving a patient is the definition of the earlier mentioned busy fraction p. As a consequence of this limitation, it makes no sense in the MCLP to locate multiple ambulances at one location. This shortcoming was addressed by Daskin (1983), by proposing the maximum expected coverage location problem (MEXCLP), which was one of the first probabilistic models for ambulance location.

In the MEXCLP, the busy fraction is incorporated as follows: if vertex $i \in V$ is covered by k ambulances, the expected covered demand is $d_i(1 - p^k)$. Moreover, the marginal contribution of the k^{th} ambulance equals $d_i(1-p)p^{k-1}$ (see also Section 4.3.1). This expression is incorporated in the objective value of the MEXCLP:

$$\text{MEXCLP: Maximize} \sum_{i \in V} \sum_{k=1}^{n} d_i (1-p) p^{k-1} z_{ik}, \tag{5.10}$$

$$\text{Subject to: } \sum_{j \in W_i} x_j \geq \sum_{k=1}^{n} z_{ik} \qquad i \in V \tag{5.11}$$

$$\sum_{j \in W} x_j \leq n \tag{5.12}$$

$$x_j \in \{0, 1, \ldots, n\} \qquad j \in W \tag{5.13}$$

$$z_{ik} \in \{0, 1\} \qquad i \in V,\ k = 1, \ldots, n. \tag{5.14}$$

Here, $z_{ik} = 1$ if and only if vertex i is covered by at least k ambulances. Note that constraint (5.12) is an inequality, while its MCLP counterpart is an equality. This is due to the concavity of the objective function in k for each i, which implies that if $z_{ik} = 1$, then $z_{i1} = z_{i2} = \ldots = z_{ik} = 1$ and if $z_{il} = 0$, then $z_{i,l+1} = z_{i,l+2} = \ldots = z_{in} = 0$. Moreover, the objective is to be maximized. Hence, constraint (5.12) will be satisfied at equality.

Analogous to the extension of the MCLP to the MEXCLP, we extend the MECRP, to address the first three shortcomings of the MECRP mentioned in Section 5.1. This is done by replacing the objective function of the MECRP, expression (5.2), by the following objective function:

$$\text{Maximize} \sum_{i \in V} \sum_{k=1}^{n} \sum_{l=1}^{k} d_i \pi_k (1-p) p^{l-1} z_{ikl},$$

where $z_{ikl} = 1$ if and only if in compliance table level k, vertex i is covered by at least l ambulances. Otherwise, $z_{ikl} = 0$. Moreover, constraint (5.3) is replaced by

$$\sum_{j \in W_i} x_{jk} \geq \sum_{l=1}^{k} z_{ikl}, \qquad i \in V,\ k = 1, \ldots, n. \tag{5.15}$$

This constraint is satisfied at equality by the same reasons as before. None of the other constraints of the MECRP change, except for constraints (5.7) and (5.8), which become $x_{jk} \in \{0, 1, \ldots, n\}$ and $z_{ikl} \in \{0, 1\}$, where $j \in W$, $i \in V$, $k = 1, \ldots, n$ and $l = 1, \ldots, k$. Moreover, constraint (5.9) is changed into $a_{jk} \in \{0, 1, \ldots, n\}$, where $j \in W$ and $k = 1, \ldots, n-1$.

5.2.3 General Performance Measures

As stated in Section 5.1, another limitation of the MECRP is the incapability to incorporate other EMS performance measures than coverage, such as patient survivability. This is a limitation of the MCLP and the MEXCLP as well. In this section we demonstrate how to incorporate different objectives in the MECRP. Similar to the previous chapters, we do this by introducing a non-negative non-decreasing penalty or cost function Φ, which is a function of the response time solely, with domain $\mathbb{R}_{\geq 0}$. A penalty function assigns to each different response time a penalty, and thus several performance measures related to response times can be incorporated. The commonly used EMS performance measure of coverage can be translated into the penalty function $\Phi(t) = \mathbb{1}_{\{t > T\}}$, where t denotes the response time and T the coverage radius, expressed in time. Other examples of objectives could be minimizing the average response time or minimizing the average lateness, modeled by penalty functions $\Phi(t) = t$ and $\Phi(t) = \max\{0, t - T\}$, respectively (see also Section 2.2.4). In addition, Erkut et al. (2008) consider survival functions, which we can use as penalty function as well (see Section 5.4).

To incorporate penalty functions and thus general performance objectives in the MECRP framework, we must be aware of the fact that coverage does not play a role here: we cannot use the set W_i defined before in our model formulation. After all, even an ambulance positioned at a location for which the travel time between this location and vertex i exceeds the coverage radius, has an effect. This effect gets larger if fewer ambulances are available. Hence, all available ambulances influence the ability to respond to a request for each vertex. In contrast, ambulances outside the coverage radius of a certain vertex i are treated as nonexistent ones for this vertex, if one uses the 0-1 nature of coverage.

As a consequence, constraint (5.3) of the MECRP needs to be replaced by a different constraint, which is able to take all available ambulances for each vertex into account. That is, for each vertex i, we need an ordering of ambulances according to their expected travel time to i, because we incorporated ambulance unavailability in our model: with probability $(1-p)$ the closest ambulance will respond to the request, generating a certain penalty $\Phi(t_1)$, with probability $(1-p)p$ the second closest ambulance will respond, generating penalty $\Phi(t_2) \geq \Phi(t_1)$, and so on, up to the k^{th} ambulance for compliance table level k. Moreover, to specify $\Phi(t_1), \Phi(t_2), \dots, \Phi(t_k)$ for compliance table level k, we need to incorporate the expected travel times $t_1, t_2, \dots, t_k$ in our model, since the penalty function relies on these.

As previously stated, the expected travel time from waiting site $j \in W$ to demand point $i \in V$ is denoted by τ_{ji}. If $\tau_{ji} \leq \tau_{j'i}$ then it holds that $\Phi(\tau_{ji}) \leq \Phi(\tau_{j'i})$ from the definition of the penalty function. Moreover, for the ordering of ambulances, we define $z_{ijkl} = 1$ if and only if for compliance table level k, the l^{th} closest ambulance to vertex i is at waiting site j. We need to introduce the constraint $\sum_{j \in W} z_{ijkl} = 1$ to ensure that at compliance table level k, there is exactly one ambulance that is the l^{th} closest to i. For an overview of the decision variables, we refer to Table 5.1. Now we have all the ingredients to formulate the minimal expected penalty relocation problem (MEXPREP).

5.2.4 Minimal Expected Penalty Relocation Problem

The MEXPREP is formulated as follows:

$$\text{Minimize} \sum_{k=1}^{n} \sum_{l=1}^{k} \sum_{i \in V} \sum_{j \in W} \pi_k d_i (1-p) p^{l-1} \Phi(\tau_{ji}) z_{ijkl} \tag{5.16}$$

$$\text{Subject to: } \sum_{l=1}^{k} z_{ijkl} = x_{jk} \qquad i \in V, j \in W, k = 1, \dots, n \tag{5.17}$$

$$\sum_{j \in W} z_{ijkl} = 1 \qquad i \in V,\ k = 1, \dots, n,\ l = 1, \dots, k \tag{5.18}$$

$$\sum_{j \in W} x_{jk} = k \qquad k = 1, \dots, n \tag{5.19}$$

$$x_{jk} - x_{j,k+1} \leq a_{jk} \qquad j \in W,\ k = 1, \dots, n-1 \tag{5.20}$$

$$\sum_{j \in W} a_{jk} \leq \alpha_k \qquad k = 1, \dots, n-1 \tag{5.21}$$

$$x_{jk} \in \{0, 1, \dots, n\} \qquad j \in W,\ k = 1, \dots, n \tag{5.22}$$

$$z_{ijkl} \in \{0, 1\} \qquad i \in V,\ j \in W,\ k = 1, \dots, n\ l = 1, \dots, k \tag{5.23}$$

$$a_{jk} \in \{0, 1, \dots, n\} \qquad j \in W,\ k = 1, \dots, n-1. \tag{5.24}$$

Note that there is only a contribution to the objective value if $z_{ijkl} = 1$, i.e., if for compliance table level k, the l^{th} closest ambulance to vertex i is at waiting

x_{jk}	Number of ambulances placed at waiting site $j \in W$ in compliance table level k.
a_{jk}	Difference in number of occurences of waiting site $j \in W$ in level k compared to level $k+1$.
z_{ijkl}	Equals 1 iff in compliance table level k, the l^{th} closest ambulance to vertex $i \in V$ is at waiting site $j \in W$, and 0 otherwise.

TABLE 5.1: Decision variables of the MEXPREP.

site j. The marginal contribution of this l^{th} closest ambulance to vertex i is $d_i(1-p)p^{l-1}\Phi(\tau_{ji})$ for given vertex i, waiting site j, and compliance table level k. That is, with probability $(1-p)p^{l-1}$, the l^{th} closest ambulance to vertex i is the closest available one, inducing a penalty of $d_i\Phi(\tau_{ji})$. Such as in the MECRP, each compliance table level k is weighted according to the probability that the system is in a situation with k available ambulances, as computed in Equation (5.1).

Constraints (5.17) and (5.18) take over the role of constraint (5.3) in the MECRP formulation. In constraint (5.17), both the left- and the right-hand side represent the number of ambulances at waiting site j for compliance table level k. Note that no i-index is present in the right-hand side. Since constraint (5.17) holds for each $i \in V$, it is immediately forced that

$$\sum_{l=1}^{k} z_{i_1jkl} = \sum_{l=1}^{k} z_{i_2jkl}, \quad i_1, i_2 \in V,\ j \in W,\ k = 1, \ldots, n.$$

This should hold in a feasible solution to the problem, since for level k all the ambulances at waiting site j contribute to the penalty induced by each demand point in the objective function. As stated before, constraint (5.18) ensures that at compliance table level k, there is exactly one ambulance that is the l^{th} closest to i. All the other constraints are the same as the constraints in the MECRP formulation, except for the integer and binary constraints. Note that since the objective is to be minimized and the penalty function $\Phi(t)$ is non-decreasing in t, we do not require constraints related to the ordering of ambulances.

5.2.5 Adjusted MEXPREP

In the MEXCLP-formulation of Daskin (1983), some simplifying assumptions are made: (1) ambulances operate independently, (2) each ambulance has the same busy fraction, and (3) ambulance busy fractions are invariant with respect to the ambulance locations. Moreover, the MEXPREP formulation, like the formulations of MEXCLP and MECRP, assumes that the busy fraction is an input. However, in reality, the busy fraction p is an output as the service rate that is needed to calculate the busy fraction depends on the allocation of ambulances to waiting sites. The use of a universal busy fraction is a rough approximation of reality, since the actual busy fractions depend on both the compliance table itself and on the dispatch policy.

Batta et al. (1989) consider an adjustment of the objective function in the MEXCLP, relaxing the assumptions on busy fractions. In this problem, called the AMEXCLP, correction factors $Q(n,p,k)$, $k=0,\ldots,n-1$, derived by Larson (1975), are incorporated in the objective function of the MEXCLP. We extend the MEXPREP to the AMEXPREP by incorporating the correction factors $Q(n,p,k-1)$ in Equation (5.16), where

$$Q(n,p,k) = \frac{\sum_{j=k}^{n-1} \frac{(n-k-1)!(n-j)}{(j-k)!} \frac{n^j}{n!} p^{j-k}}{(1-p)\sum_{i=0}^{n-1} \frac{n^i}{i!} p^i + \frac{n^n p^n}{n!}}, \quad k=0,\ldots,n-1, \tag{5.25}$$

analogous to the work done by Batta et al. (1989). In Section 5.4.6, we explore the differences between the MEXPREP and the AMEXPREP.

5.3 Assignment Problem

Determining the compliance table is just the first part of the ambulance relocation problem. The second part is related to the actual assignment of the k available ambulances to the k waiting sites occuring in compliance table level k. This problem is studied extensively by Maleki et al. (2014), and two models for determining the assignment of ambulances to the waiting sites in compliance table level k, as computed via solving the MECRP, are proposed. In each of these two models, called the generalized ambulance assignment problem (GAAP) and the generalized ambulance bottleneck assignment problem (GABAP), a different, yet related, objective is incorporated: GAAP minimizes the total travel time traveled by all ambulances to attain the configuration of the compliance table level, while GABAP minimizes the maximum travel time. Both, like the MECRP, are offline methods, computing assignments beforehand. However, scalability issues are present, since the number of combinations between hospitals/waiting sites and waiting sites grows very rapidly.

As opposed to the offline approach of Maleki et al. (2014), we use an online approach in our computations, by modeling the assignment problem as either a minimum weighted bipartite matching problem (MWBM) or a linear bottleneck assignment problem (LBAP). By modeling the problem as a MWBM, we aim to find an assignment of available ambulances to the designated waiting sites in the compliance table that minimizes the total travel time. However, in the assignment, it may happen that one ambulance needs to make a very long trip. Hence, the area around the waiting site to which this ambulance is assigned is vulnerable for a long time. It may be advantageous to minimize the maximum travel time, and thus the time until the system is in compliance. This can be done by modeling the assignment problem as an LBAP.

In contrast to the computation of compliance tables, fast methods exist for solving the MWBM and the LBAP, e.g., the Hungarian Method of complexity $\mathcal{O}(n^3)$ for MWBM and the Threshold Algorithm of complexity $\mathcal{O}(n^{2.5}/\sqrt{\log n})$ for LBAP, both explained by Burkhard et al. (2009). Hence, this can be done in real-time and an offline solution is not necessary. After all, this would require a complex state dependent policy which shows relocation moves for every realized state of the system. Moreover, an online implementation of the assignment problem allows takes into account the actual locations of driving ambulances and hence a redirection of ambulances to different waiting sites. Therefore, we recommend to compute compliance tables offline, and the assignment problem online. In Section 5.4.4, we will explore the differences in the MWBM and the LBAP.

5.4 Computational Study

The MEXPREP computes compliance tables taking into account ambulance unavailability, general performance measures, and a restriction on the number of waiting site changes. We apply the MEXPREP to the Amsterdam EMS region, extensively described in Section 3.4.2. In this chapter, we assume that ambulances can idle at any of the 17 dots depicted in Figure 3.4. Results are generated by simulation using historical data.

5.4.1 Experimental Setup

Historical data on emergency requests in the year 2011 was provided by Ambulance Amsterdam, which runs the emergency medical services in this region. We only consider the time-period between 7 AM and 6 PM, like in Section 4.4.1. During the considered time-period, 33 ambulances are present in the system. However, of these ambulances, many are busy with ordered transport: taxi-type transport of patients not able to travel to the hospital themselves, usually scheduled in advance. Therefore, we assume a fleet size of 21 in our computations.

In 2011 between 7 AM and 6 PM, the total number of emergency requests was 44,966, yielding an hourly arrival rate of 11.2 requests. Only 44,520 of these requests are useful, because of the remainder historical data was not complete. We build a trace on this data and simulate it in a discrete-event simulation. We refer to Section 4.4.1 for an enumeration of the incident related information included in the trace. Unlike the mentioned section, we do not remove days, as the fleet capacity of 21 ambulances is satisfactory. We connect the 365 days in the trace such that 6 PM is followed directly by 7 AM, for the same reason as explained in Section 4.4.1.

We also use historical data to compute the busy fraction, by dividing the total patient-related work during these 4,015 hours by the total duty time of 21 ambulances, to obtain a busy fraction of $p = 0.43047$. The average busy time (excluding relocation time after transferring the patient at the hospital) of an ambulance is 0.82 hours. The annual number of emergency requests ranges between 2 (in a postal code somewhere between waiting sites 9 and 13) and 1,545 (in the city center of Amsterdam, near waiting site 1), with an average demand of 275 per node. We define d_i as the probability that an incoming request occurs in vertex i, computed by normalization of the number of emergency requests.

We assume a deterministic dispatch time of 120 seconds and a deterministic chute time of 60 seconds for ambulances at a waiting site. There is no chute time if the dispatched ambulance is already on the road. Moreover, the pre-trip delay for moving an ambulance from a waiting site to another one is assumed to be 180 seconds. The ambulance that can be present fastest at the emergency scene is always dispatched to the request.

We perform simulations using the computed compliance tables and the actual emergency requests in the region during the daytime of the year 2011. To keep track of the actual locations of ambulances, we use the travel routes as computed in Section 3.4.3, which use the travel time table estimated by the RIVM (Kommer and Zwakhals, 2008) as input. The relocation travel times were computed by multiplying the emergency travel times by a factor $\frac{10}{9}$. Computation of the assignment of ambulances to waiting sites is done online by solving either the MWBM or the LBAP during the simulation. We test performance according to six statistics:

1. Percentage requests responded to within the response time threshold (720 seconds).

2. Average penalty per request.

3. Average response time.

4. Average number of relocations per ambulance per day. A move of an ambulance only counts as relocation if this move is induced by carrying out the compliance table policy.

5. Average relocation time.

6. Computation time to solve the model, run with CPLEX 12.6 on a 2.2 GHz Intel(R) Core(TM) i7-3632QM laptop with 8 GB of RAM.

In our computations, we consider five different penalty functions. Three of them are based on survival functions, considered by De Maio et al. (2003), by Valenzuela et al. (1997), and by Waaelwijn et al. (2001). These three functions all relate a survival probability to a response time, in the case of a cardiac arrest. However, these survival probabilities depend on additional factors rather than just the response time, e.g., whether the collapse of a patient was witnessed by the ambulance crew, the duration from collapse to defibrillation, and the duration from collapse to cardiopulmonary resuscitation (CPR). These three survival functions are considered by Erkut et al. (2008), and assumptions on these factors are made. We follow these assumptions to obtain a survival function solely depending on the response time (in seconds). The considered penalty functions are as follows:

$$\Phi_1(t) = \mathbb{1}_{\{t>720\}}, \tag{5.26}$$
$$\Phi_2(t) = t, \tag{5.27}$$
$$\Phi_3(t) = 1 - (1 + e^{0.679+0.0044t})^{-1}, \tag{5.28}$$
$$\Phi_4(t) = 1 - (1 + e^{0.113+0.0041t})^{-1}, \tag{5.29}$$
$$\Phi_5(t) = 1 - (1 + e^{0.04+0.005t})^{-1}. \tag{5.30}$$

Function Φ_1 is based on coverage, in which we consider a response time threshold of 720 seconds (12 minutes). Φ_2 represents the penalty function focusing on the

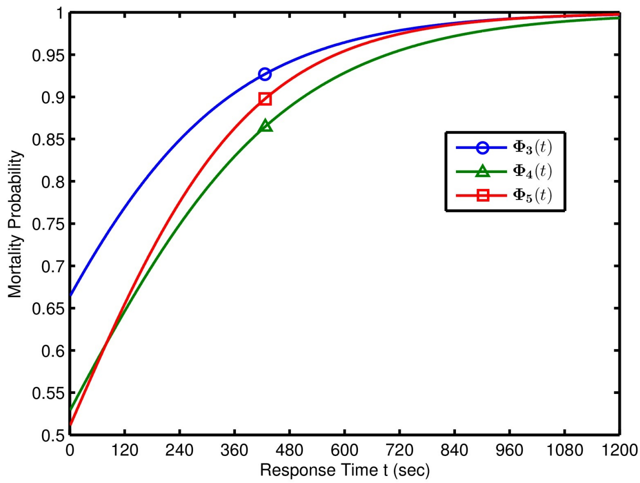

Figure 5.1: Mortality probabilities as a function of the response time.

objective of minimizing the average response time. Functions Φ_3, Φ_4, and Φ_5 represent the survival functions of De Maio et al. (2003), Valenzuela et al. (1997), and Waaelwijn et al. (2001), respectively, in a penalty function (mortality) setting. A graphical representation of Φ_3, Φ_4, and Φ_5 is given in Figure 5.1.

5.4.2 Comparison of MEXPREP with MECRP

First, we compare the compliance tables obtained by MEXPREP with the ones obtained by MECRP, following the formulation proposed by Gendreau et al. (2006). We do this for the coverage-based penalty function $\Phi(t) = \mathbb{1}_{\{t>r\}}$, since the MECRP cannot take other penalty functions into account. We use a coverage radius of $T = 720$ seconds (12 minutes), and compute compliance tables for different values of α_k. Due to the incapability of the MECRP to consider systems with more ambulances than waiting sites, which is the case here, we compare the MEXPREP with the MECRP on two different settings: a setting with 17 ambulances instead of 21; and a setting in which we have 21 ambulances, but the compliance table will be carried out only if 17 or fewer ambulances are available. If more than 17 ambulances are available, ambulances that finish service of a patient return to their home waiting site. In the first setting, the busy fraction is 0.53175, while in the second setting the busy fraction equals 0.43047 as mentioned before.

We only display the compliance tables for the $\alpha_k = 0$ case, since these compliance tables are nested and thus can be represented efficiently. We represent such a nested compliance table by a one-dimensional vector, where compliance table level k is given by entries 1 up to k. The computed MECRP and MEXPREP compliance tables for $\alpha_k = 0$, for the two different settings are displayed in Equations (5.31) and (5.32), respectively. Note that none of these four compliance tables equals an-

Method	Performance Indicators	$\alpha_k = 0$	$\alpha_k = 1$	$\alpha_k = \lceil \frac{k}{2} \rceil$	$\alpha_k = k$
MECRP	Percentage on time	86.55%	86.29%	86.62%	86.60%
	Lower Bound 95%-CI	86.24%	85.97%	86.31%	86.28%
	Upper Bound 95%-CI	86.87%	86.60%	86.94%	86.92%
	Mean response time	473 s	476 s	474 s	474 s
	Mean no. relocations	1.62	2.14	3.86	3.72
	Mean relocation time	646 s	576 s	451 s	457 s
	Computation time	<1 s	<1 s	<1 s	<1 s
MEXPREP	Percentage on time	88.23%	88.18%	88.18%	88.34%
	Lower Bound 95%-CI	87.93%	87.88%	87.88%	88.04%
	Upper Bound 95%-CI	88.53%	88.48%	88.48%	88.64%
	Mean response time	461 s	461 s	461 s	460 s
	Mean no. relocations	1.30	1.31	1.31	1.54
	Mean relocation time	625 s	616 s	616 s	571 s
	Computation time	76 s	85 s	85 s	77 s

Table 5.2: Simulation results for 17 ambulances and penalty function $\Phi(t) = \mathbb{1}_{\{t>720\}}$, based on 44,520 requests in 2011.

other, although the two MECRP-tables are very similar. Simulation results, using MWBM as assignment policy, for these compliances tables are listed in Tables 5.2 and 5.3, respectively. These tables include 95% confidence intervals around the percentage of requests responded to within 720 seconds.

$$\begin{aligned} \text{MECRP:} &\quad (1, 16, 12, 14, 5, 9, 17, 8, 11, 15, 3, 10, 4, 13, 2, 6, 7) \\ \text{MEXPREP:} &\quad (1, 1, 6, 16, 6, 15, 2, 10, 16, 14, 1, 10, 15, 9, 6, 17, 12) \end{aligned} \tag{5.31}$$

$$\begin{aligned} \text{MECRP:} &\quad (1, 16, 12, 14, 5, 9, 17, 8, 10, 15, 11, 3, 4, 13, 2, 6, 7) \\ \text{MEXPREP:} &\quad (1, 6, 16, 1, 15, 10, 2, 14, 6, 16, 10, 9, 17, 2, 12, 14, 5) \end{aligned} \tag{5.32}$$

Note that in Table 5.2 as well as in Table 5.3, the MEXPREP significantly outperforms the MECRP on the most important performance indicator: the percentage of requests responded to within the response time threshold of 720 seconds. We observe improvements on this criterion between 0.7% (second setting, $\alpha_k = 0$) and 1.89% (first setting, $\alpha_k = 1$). Moreover, this performance gain is achieved with fewer relocations, although the average relocation time is longer for MEXPREP. A small disadvantage of the MEXPREP compared to the MECRP is the computation time. However, as stated before, the computation time of the MEXPREP compliance tables is of less importance, since the problem can be solved in an offline fashion.

Observing the results listed in Tables 5.2 and 5.3, we note that the benefit of allowing non-nested compliance tables is very marginal with respect to the percentage of requests for which the response time threshold is achieved, and to the average response time. In some cases it is even disadvantageous to allow more

Method	Performance Indicators	$\alpha_k = 0$	$\alpha_k = 1$	$\alpha_k = \lceil \frac{k}{2} \rceil$	$\alpha_k = k$
MECRP	Percentage on time	94.39%	94.27%	94.11%	94.09%
	Lower Bound 95%-CI	94.17%	94.06%	93.89%	93.87%
	Upper Bound 95%-CI	94.60%	94.49%	94.33%	94.31%
	Mean response time	415 s	417 s	418 s	418 s
	Mean no. relocations	2.64	3.79	4.16	4.17
	Mean relocation time	509 s	444 s	420 s	420 s
	Computation time	< 1 s	< 1 s	< 1 s	< 1 s
MEXPREP	Percentage on time	95.09%	95.09%	95.09%	95.17%
	Lower Bound 95%-CI	94.89%	94.89%	94.89%	94.97%
	Upper Bound 95%-CI	95.30%	95.30%	95.30%	95.37%
	Mean response time	416 s	416 s	416 s	412 s
	Mean no. relocations	1.53	1.53	1.53	2.88
	Mean relocation time	675 s	675 s	675 s	515 s
	Computation time	67 s	67 s	67 s	72 s

TABLE 5.3: Simulation results for 21 ambulances, compliance tables up to level 17 and penalty function $\Phi(t) = \mathbb{1}_{\{t>720\}}$, based on 44,520 requests in 2011.

than zero waiting site changes. Besides that, in the second setting the MEXPREP computes the same compliance tables for $\alpha_k = 0$, $\alpha_k = 1$ and $\alpha_k = \lceil \frac{k}{2} \rceil$. However, the effect on the number of relocations is large if one uses the compliance tables with no restrictions on waiting site changes rather than compliance tables with restrictions. The question arises whether this marginal performance improvement outweighs this increase in number of relocations. In line with Gendreau et al. (2006), the average relocation time decreases if more waiting site changes are allowed, as expected.

5.4.3 Relocation Thresholds

The number of relocations in Table 5.3 is quite large. For instance, for the MEXPREP with $\alpha_k = 0$, the average number of relocations per day is 32. This is due to the large number of changes in availability of ambulances. After all, each time an ambulance is dispatched or finishes service, relocations may be performed. However, one could argue the effect of ambulance relocations if enough ambulances are still available. As example, it probably makes no sense to relocate ambulances if $n - 1$ instead of n ambulances are available, since frequent movements may inconvenience ambulance crews. A way to address this is the introduction of a *relocation threshold*, denoted by K. If the number of available ambulances is below this threshold, we use the compliance table policy. However, if this is not the case, we carry out the *static policy*: we perform no relocations if an ambulance is dispatched, and we send a newly finished ambulance back to its home waiting site. If a transition from level K to $K + 1$ occurs, each ambulance is sent back to its home waiting site. Note that these ambulance movements do *not* contribute to the number of relocations, as it is beneficial from the crew's perspective to be present at the home waiting site.

The determination of the ideal level of this relocation threshold K is an interesting topic. If K is too high, it is possible that too many relocations are performed. On the other hand, a low value of K may result in a worse performance of an ambulance service provider. To investigate the behavior of different relocation thresholds K, we compute compliance tables by the MEXPREP for $K = 7$, $K = 14$, and $K = 21$, for the five different penalty functions of (5.26)–(5.30), where $\alpha_k = 0$. That is, we change n in the MEXPREP-formulation to K and compute K compliance table levels. Except for the fact we do not change the π_k-values in the objective function, we compute the MEXPREP as if there were K ambulances instead of n.

In addition, we compute an initial configuration of the $n = 21$ ambulances by an ordinary location problem, which is a modification of the MEXPREP, as follows. In the MEXPREP, we set $k = 21$ in all constraints and in the objective function. Moreover, we discard constraints (5.20), (5.21) and (5.24), as well as π_k in the objective function. Note that for penalty function Φ_1 this modification of the MEXPREP is equivalent to the MEXCLP.

Then, we simulate our system for $K = 0$ (the static policy), $K = 7$, $K = 14$ and $K = 21$, starting in the initial configuration. This initial configuration also determines the home waiting site of each ambulance. In the simulation, we solve the MWBM to obtain a solution to the assignment problem. The results are listed in Table 5.4.

As expected, the performance on the patient-based performance indicators (which are fractions on time, average penalty, and average response time) increase as K increases. Specifically, the compliance tables obtained by the MEXPREP outperform the static policies, which in addition to the MECRP compliance tables could also serve as a benchmark policy on all penalty functions. However, this comes at the expense of additional ambulance relocations.

Interestingly, fewer ambulance relocations are performed when a relocation threshold $K = 21$ is used instead of $K = 14$. This behavior is easily explained by the following observation: the majority of the ambulance relocations are done when a transition from level $K + 1$ to K occurs. If $K = n = 21$, there are no transitions from level $K+1$ to level K. Due to the nesting of the compliance table, relatively few ambulance relocations are performed. However, for $K = 14$, there are many transitions from level 15 to level 14. Together with the fact that level 14 is generally not nested in the ambulance configuration with 15 ambulances, many ambulance relocations are carried out. This behavior is also reflected in Figure 5.2, where the total number of relocations and mean penalty as a function of K is displayed. It is not a surprise that the peak of the number of relocations is at $K = 12$. After all, the mean number of available ambulances is between 12 and 13, so many transitions from a situation with 13 to a situation with 12 available ambulances take place.

Note that for the static policy $K = 0$, the performance indicators differ for the considered penalty functions in general, although no compliance table policy is carried out. This is a direct consequence of the differences in the initial con-

Function	Performance Indicators	$K = 0$	$K = 7$	$K = 14$	$K = 21$
Φ_1	Percentage on time	90.41%	91.36%	95.02%	95.19%
	Average penalty	0.10	0.09	0.05	0.04
	Mean response time	462 s	456 s	422 s	415 s
	Mean no. relocations	0	1.09	2.07	1.40
	Mean relocation time	-	707 s	676 s	678 s
	Computation time	-	5 s	38 s	470 s
Φ_2	Percentage on time	93.54%	94.09%	95.47%	95.66%
	Average penalty	433	429	405	403
	Mean response time	433 s	429 s	405 s	403 s
	Mean no. relocations	0	1.13	2.30	1.60
	Mean relocation time	-	604 s	595 s	647 s
	Computation time	-	6 s	35 s	172 s
Φ_3	Percentage on time	93.26%	93.92%	95.08%	95.13%
	Average penalty	0.9124	0.9114	0.9052	0.9043
	Mean response time	431 s	426 s	405 s	402 s
	Mean no. relocations	0	1.02	2.41	1.75
	Mean relocation time	-	632 s	574 s	603 s
	Computation time	-	7 s	68 s	455 s
Φ_4	Percentage on time	93.26%	93.89%	95.08%	95.11%
	Average penalty	0.8464	0.8447	0.8351	0.8341
	Mean response time	431 s	426 s	405 s	403 s
	Mean no. relocations	0	1.02	2.41	1.76
	Mean relocation time	-	633 s	574 s	608 s
	Computation time	-	5 s	50 s	548 s
Φ_5	Percentage on time	93.26%	93.89%	95.05%	95.09%
	Average penalty	0.8741	0.8726	0.8632	0.8614
	Mean response time	431 s	426 s	405 s	402 s
	Mean no. relocations	0	1.02	2.39	1.78
	Mean relocation time	-	633 s	575 s	600 s
	Computation time	-	4 s	107 s	713 s

Table 5.4: Simulation results for several levels of K, $n = 21$ ambulances and $\alpha_k = 0$, based on 44,520 requests in 2011.

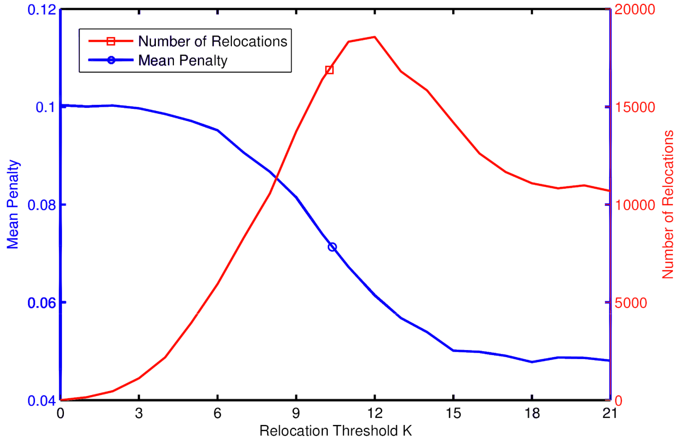

Figure 5.2: Total number of relocations and mean penalty as a function of the relocation threshold K, for 21 ambulances, $\alpha_k = 0$ and penalty function $\Phi(t) = \mathbb{1}_{\{t>720\}}$.

figurations. Moreover, it is worth noting that the coverage penalty function Φ_1 is outperformed by the average response time penalty function Φ_2 on the percentage on time criterion, despite the fact that Φ_1 focuses on maximizing this percentage. This underlines the conclusion made by Erkut et al. (2008) about the weakness of models based on coverage.

5.4.4 Assignments

We proceed this numerical study with a comparison of the two models for solving the assignment problem, mentioned in Section 5.3, namely the MWBM and the LBAP.

The results in Table 5.5 show that using the LBAP for the assignment problem results in a slightly better performance regarding the patient-based performance indicators. This small increase is explained by the observation that the LBAP minimizes the maximum travel time of a relocated ambulance. As a consequence, the ambulance configuration corresponding to the new compliance table level is attained faster. Hence, as expected, the average relocation time per ambulance decreases drastically. After all, using the LBAP, a long trip of one ambulance is split into multiple shorter trips, thus reducing the average relocation time per ambulance. However, the total number of relocations is approximately quadrupled with respect to the usage of the MWBM as assignment problem. This is probably not acceptable from the crew perspective. It is up to the ambulance service provider to decide whether this tremendous increase of number of relocations outweighs the benefits of the increase in patient-based performance.

Method	Performance Indicators	Φ_1	Φ_2	Φ_3	Φ_4	Φ_5
MWBM	Percentage on time	95.19%	95.66%	95.13%	95.11%	95.09%
	Average penalty	0.04	403	0.9043	0.8341	0.8614
	Mean response time	415 s	403 s	402 s	403 s	402 s
	Mean no. relocations	1.40	1.60	1.75	1.76	1.78
	Mean relocation time	678 s	647 s	603 s	608 s	600 s
	Computation time	470 s	172 s	455 s	548 s	713 s
LBAP	Percentage on time	95.62%	95.71%	95.23%	95.26%	95.27%
	Average penalty	0.04	394	0.9017	0.8300	0.8579
	Mean response time	408 s	394 s	395 s	395 s	396 s
	Mean no. relocations	6.11	6.33	6.47	6.52	6.51
	Mean relocation time	394 s	387 s	365 s	363 s	361 s
	Computation time	467 s	172 s	440 s	552 s	710 s

TABLE 5.5: Simulation results for $n = K = 21$ and $\alpha_k = 0$, based on 44,520 requests in 2011.

Evaluation:	Φ_3		Φ_4		Φ_5	
	MWBM	LBAP	MWBM	LBAP	MWBM	LBAP
Φ_1	4,033	4,163	7,056	7,248	5,803	6,003
Φ_2	4,228	4,350	7,355	7,537	6,106	6,294
Φ_3	4,261	4,378	7,404	7,577	6,159	6,339
Φ_4	4,250	4,372	7,387	7,567	6,142	6,329
Φ_5	4,268	4,371	7,413	7,565	6,170	6,328

TABLE 5.6: Expected number of survivors for $n = 21$ and $\alpha_k = 0$, based on 44,520 requests in 2011.

5.4.5 Expected Number of Survivors

Another interesting indicator that provides insight into the performance of the compliance tables, is the expected number of survivors. This expected number is easily computed by the summation of the 44,520 penalties for the survival functions Φ_3, Φ_4 and Φ_5. Moreover, we perform cross-comparisons of these functions: we evaluate the compliance table corresponding to the solution of the MEXPREP for one specific penalty function (rows) using the other ones (columns), for both the MWBM and the LBAP. The results are listed in Table 5.6.

If one considers the rows corresponding to Φ_3, Φ_4 and Φ_5 in Table 5.6, one may observe that the differences within these columns are small: the numbers differ at most by 0.5%. We conclude that the chosen survival function is not of influence on the maximization of survivors. In contrast, the number of survivors differs for the compliance tables induced by the penalty functions based on coverage and average response times, Φ_1 and Φ_2, respectively. Especially for Φ_1, this difference

	Performance Indicators	Φ_1	Φ_2	Φ_3	Φ_4	Φ_5
$\alpha_k = 0$	MEXPREP Objective value	0.0572	443	0.9172	0.8533	0.8817
	MWBM simulated penalty	0.0438	403	0.9043	0.8341	0.8614
	LBAP simulated penalty	0.0438	395	0.9012	0.8293	0.8573
$\alpha_k = k$	MEXPREP Objective value	0.0571	443	0.9172	0.8533	0.8817
	MWBM simulated penalty	0.0439	403	0.9038	0.8328	0.8608
	LBAP simulated penalty	0.0426	396	0.9013	0.8292	0.8566

TABLE 5.7: MEXPREP objective values and simulated penalties for $n = 21$, based on 44,520 requests in 2011.

is around 5% compared to the survival functions. However, the difference between the survival functions and Φ_2 is relatively minor. As a consequence, it seems that the average response time is a better approximation for survival than coverage.

As can be observed in Table 5.6, there are differences between the MWBM and the LBAP. For instance, the expected number of survivors using the LBAP increases with approximately 2.6% with respect to the case in which the MWBM is used as assignment problem, for Φ_3. This was to be expected due to the increase in performance of the LBAP with respect to the MWBM, as can be observed in Table 5.5. The expected number of survivors is smallest when the compliance tables are evaluated using penalty function Φ_3. This is explained by the fact that Φ_3 is the most pessimistic survival function (see Figure 5.1).

5.4.6 AMEXPREP

In Section 5.2.5, we discussed some limitations and assumptions on busy fractions. These assumptions may result in a objective value of the MEXPREP that differs from the values computed through simulation. In Table 5.7, objective and simulated values are listed for the two extremes $\alpha_k = 0$ and $\alpha_k = k$, for both the MWBM and the LBAP.

From Table 5.7, we conclude that MEXPREP's estimation of the system performance is somewhat too pessimistic. This is most evident in Φ_1, in which the relative gap between objective value and simulated values is largest. Moreover, we observe a difference only in the fourth digit in the objective values for $\alpha_k = 0$ and $\alpha_k = k$ for Φ_1. From this observation, one could draw the conclusion that nested compliance tables are already close to optimal. This is also underlined by the simulated values. In all cases, the simulated values using the MWBM are closer to the objective values than in the simulation that uses the LBAP as the assignment problem. This is as expected, since the use of the LBAP results in better patient-based performance (see Table 5.5).

As opposed to the objective values of the MEXPREP, the AMEXPREP presented in Section 5.2.5 provides an optimistic estimation of the system performance, as can be observed in Table 5.8. For the penalty functions based on survival, Φ_3, Φ_4 and Φ_5, the objective value of the AMEXPREP differs more from

	Performance Indicators	Φ_1	Φ_2	Φ_3	Φ_4	Φ_5
MEXPREP	Objective value	0.0572	443	0.9172	0.8533	0.8817
	MWBM simulated penalty	0.0438	403	0.9043	0.8341	0.8614
	LBAP simulated penalty	0.0438	394	0.9017	0.8300	0.8579
AMEXPREP	Objective value	0.0371	380	0.8127	0.7539	0.7794
	MWBM simulated penalty	0.0435	400	0.9032	0.8323	0.8600
	LBAP simulated penalty	0.0423	395	0.9014	0.8293	0.8574

TABLE 5.8: AMEXPREP objective values and simulated penalties for $n = 21$ and $\alpha_k = 0$, based on 44,520 requests in 2011.

	Φ_1		Φ_2		Φ_3		Φ_4		Φ_5	
	c1	*c2*	*c1*	*c2*	*c1*	*c2*	*c1*	*c2*	*c1*	*c2*
Deviations	13	16	4	4	5	9	5	10	11	14
Levels	9-21	9-21	18-21	18-21	17-21	2,3,18-21	17-21	2-4,18-21	17-21	2-4,18-21

TABLE 5.9: Deviations of unrestricted MEXPREP compliance tables with respect to actual capacities and restricted MEXPREP for $\alpha_k = 0$.

the simulated values than is the case for the MEXPREP. Surprisingly, for Φ_1 and Φ_2 it is the opposite. At last, it is worth noting that the AMEXPREP performs slightly better than the MEXPREP on the penalty criterion in general.

5.4.7 Base Station Capacities

In this section, we solve the MEXPREP, taking into account the actual waiting site capacities depicted in Figure 3.4. These restrictions can easily be incorporated in the MEXPREP by introducing constraints of the type

$$x_{jk} \leq c_j \qquad j \in W,\ k = 1, \dots, n-1, \tag{5.33}$$

where c_j denotes the capacity of waiting site $j \in W$. We compute the restricted version of MEXPREP for $\alpha_k = 0$. We compare the obtained compliance table to the actual capacities. The number of deviations is reported in the columns *c1* in Table 5.9. For instance, for Φ_1, the number of capacity violations is 13 for the whole compliance table, and these violations occur in levels nine up to 21. In addition, columns *c2* report the numbers for the restricted compliance table compared to the unrestricted one. Note that the compliance tables consist of 231 numbers in total.

Only for Φ_1 the computation of the restricted MEXPREP results in a different objective value compared to the unrestricted MEXPREP: 0.0576. For the other penalty functions, the objective values do not differ in the first four digits, although different compliance tables were generated, as can be observed in Table 5.9. From this observation, one could draw the conclusion that minor differences in compliance tables are hardly noticed in the objective value: there are many compliance

$\|V\|$	100	200	300	400	500	600
Number of variables	3.2×10^5	6.3×10^5	9.5×10^5	1.3×10^6	1.6×10^6	1.9×10^6
Number of constraints	3.1×10^5	6.2×10^5	9.2×10^5	1.2×10^6	1.5×10^6	1.8×10^6
CPU time $\Phi_2,\ \alpha_k = 0$	53 s	168 s	348 s	695 s	1119 s	1770 s
CPU time $\Phi_5,\ \alpha_k = 0$	38 s	196 s	387 s	808 s	1182 s	1689 s
CPU time $\Phi_2,\ \alpha_k = k$	63 s	197 s	459 s	595 s	1049 s	1594 s
CPU time $\Phi_5,\ \alpha_k = k$	47 s	189 s	349 s	576 s	997 s	1650 s

TABLE 5.10: Computation times for the artificial problem instance.

tables that are near-optimal. It is also interesting to see that there are deviations in lower levels for penalty functions Φ_3, Φ_4 and Φ_5 with respect to the restricted compliance table, while these are not present in the middle levels.

In addition, we simulate the restricted compliance tables. The differences in average penalties between restricted and unrestricted compliance tables are very small for all penalty functions and not worth reporting. According to this analysis, one might conclude that the current capacity is not a limiting factor.

5.4.8 Computation Times

We conclude this section with an investigation on computation times of the MEXPREP. Unfortunately, we are not able to investigate the increase in computation time by choosing a different demand aggregation for the considered case, since we only have access to travel times between 4-digit postal codes. As an alternative, we create an artificial problem instance: we pick $|V|$ demand nodes out of a grid of size 100×100, for different values of $|V|$, and assign demand probabilities to them. Travel times between nodes are calculated by the Manhattan metric. For the base locations, we select $|W| = 15$ points, and we consider $n = 20$ ambulances. Then, we solve the MEXPREP for the extremes $\alpha_k = 0$ and $\alpha_k = k$, and for Φ_2 and Φ_5, since in Table 5.5 the computation time of these penalty functions is shortest and longest, respectively. Results on computation times, as well as number of variables and constraints (namely, Equations (5.17)–(5.21)) are listed in Table 5.10.

For large values of $|V|$, it takes more time to obtain a solution for $\alpha_k = 0$ compared to $\alpha_k = k$, as can be observed in Table 5.10. The explanation of this phenomenon is probably in the method CPLEX uses to compute a solution. From Tables 5.4 and 5.5 one may conclude that the use of Φ_2 and Φ_5 induce the shortest and longest computation times, respectively. However, Table 5.10 shows that Φ_2 did not consistently result in shorter computation times than Φ_5.

5.5 Concluding Remarks

In this paper, we presented the minimum expected penalty relocation problem (MEXPREP) to compute compliance tables. The MEXPREP is an extension of the maximal covering relocation problem (MECRP) formulated by Gendreau et al. (2006) in two directions. First, we incorporated the objective function of the

MEXCLP into the objective function of the MECRP, to anticipate multiple future emergency requests beyond a first request. Then, we introduced penalty functions in order to focus on performance measures other than coverage, including survival probabilities. Moreover, based on the assumptions and limitations of busy fractions, we introduced an adjusted version of the MEXPREP. In this adjusted version, called the AMEXPREP, correction factors proposed by Batta et al. (1989) were incorporated. Additionally, we considered both the minimum weighted bipartite matching problem (MWBM) and the linear bottleneck assignment problem (LBAP) as assignment problem for the assignment of available ambulances to the waiting sites indicated by the compliance table level.

We concluded this paper with a numerical study, based on 44,520 emergency requests in 2011 in the region of Amsterdam and its surroundings. In this study, we compared the MEXPREP compliance tables to both the MECRP compliance tables and the static policy, and we observed that the MEXPREP outperforms both of them on most performance indicators. We also carried out a comparison between several restrictions on waiting site changes. Moreover, we considered several relocation thresholds, and compared both the resulting performance when using the LBAP and the MWBM as assignment problems. In addition, we compared the objective values with the simulated values for both the MEXPREP and the AMEXPREP. Studies regarding computation times of the MEXPREP and the effect of base station capacities were conducted as well.

There are several extensions that can be made to improve the realism of the MEXPREP model. For instance, we assumed travel times to be deterministic, while in reality these are stochastic. Moreover, we used one universal busy fraction p, which induce some limitations. For instance, in reality, this busy fraction probably differs per base location. Another interesting research topic is a modification of the MEXPREP in which only certain designated levels of the compliance table are computed, rather than the whole compliance table, and how this kind of policy effects the performance. With regard to survival probabilities, we only considered survival functions based on a cardiac arrest, while other types of emergency requests occur in practice as well. However, survival functions for several types of emergency requests could be combined in one survival function using weights corresponding to the frequency of different request types (if this could be quantified, as pointed out by Erkut et al. (2008)). The MEXPREP model to compute compliance tables presented in this paper forms a good basis for these extensions and modifications.