On-route coverage by available ambulances through route optimization

Martin van Buuren, Coen W. P. Huibers, and Rob D. van der Mei

Abstract

Dispatchers usually relocate ambulances by the fastest route. However, there are cases when the fastest relocation is not the best one, as the transitory coverage while driving over the fastest route can be relatively low. When a fast relocation happens over the highway with a limited number of exits the ambulance is less flexible, while a longer route through villages can lead to a higher performance due to transitory coverage. In this paper, we study the effect of taking alternative routes in ambulance relocations, which is an unaddressed problem in the literature. Results for four actual ambulance regions show that this so-called *dynamic routing* can help at an operational level to obtain a more fair distribution of ambulance coverage.

1 Introduction

Dynamic ambulance management, DAM, distributes available ambulances over an ambulance region to minimize late arrivals. When an ambulance finishes a service at a hospital, or when a dispatch to a new incident occurs, it is generally beneficial to ask available ambulances to drive to another location than a fixed base location, such that the coverage increases. An ambulance movement for improved coverage is called a *relocation*. Relocations may only go to a limited number of predefined relocation points—often these are base locations.

The DAM models known from literature often start by determining for each ambulance to what base locations it can be relocated, and consequently, what configuration scores best. After calculating the optimal configuration, the output for each ambulance from what origin to what destination it must move; this is called an *OD-pair*. These models, however, do not specify *how* an ambulance should drive to the destination. Usually it is assumed that the ambulances take the fastest route.

Ambulances in densely populated areas such as the Netherlands are distributed for fairness: no matter where in the country you are, there must be a reasonable probability that an ambulance is nearby. Because the performance indicators are aggregated over the year and the entire ambulance region, there is an incentive to concentrate ambulances around cities with high demand. Paper [2] and Ph.D. thesis chapters [1] address this issue and propose facility and allocation models to increase fairness.

The main contribution of the current chapter is that we focus on the question *what route* the ambulance should take to drive to its destination. While an available ambulance is moving to its destination, the route choice has an influence on the coverage, and thereby, both on the volume and the spatial distribution of late arrivals. This so-called *dynamic routing* can be highly effective in increasing fairness over an ambulance region, particularly for regions where subareas are not covered from existing base locations. We show results for three ambulance regions.

A literature review is provided in Section 2. Next, in Section 3, we propose our dynamic routing model. Results for actual ambulance regions are provided in Section 4. This chapter ends with a conclusion in Section 5.

2 Literature

Many methods address the issue of *how* to choose the ambulance movement, see Section **??** for a literature overview on this topic. We first discuss ways for alternative routing. Directly after, we give a brief overview of the *Dynamic Maximum Expectated Coverage Location Problem* (DMEXCLP) that is used in this chapter.

Generating a routeset, i.e., a set containing route alternatives, in combination with a route choice model is well studied in the literature. An early review of this topic is given in [19]. Further extensive literature can be found in [21] and [22].

Various techniques are proposed to deal with uncertainty in travel costs [9, 18, 13]. Stochastic approaches calculate iteratively each route alternative in two phases. In the first phase, the edge costs are drawn from a distribution, and in the second phase a shortest path is calculated [11, 7]. Doubly stochastic models are an extension that also randomize the objective function, in the case it consists of multiple weighted link properties [8, 15]. The labeling approach generates a routeset for multi-criteria objective by including one route for each of the criteria [6]. Link elimination alternatively executes two procedures: a shortest path algorithm and a path deletion algorithm that deletes the characteristic link [20, 4]. A breath-first search can be appended to link elimination to speed up the process of generating a high diversity of paths [22].

The models mentioned above are designed to find a good fastest route. Our objective, however, differs because we wish to incorporate the coverage provided *by a route*.

The route selection problem for hazardous material transports is a relatable problem to ambulance routing, since both involve coverage. Instead of taking routes with the *least* coverage, the ambulance context wants us to take a route that generates the *most* coverage in their objective function. A known approach is to first define the edge risk for each road segment (that is the probability that an incident occurs on an edge multiplied by the damage), and thereafter to calculate a minimal path [23, 12, 3, 24].

There are differences between hazardous materials and real-time ambulance allocation is the required calculation speed. Whereas hazardous material routes can be calculated weeks in advance, the ambulance dispatch centers need a faster method that can provide an answer within seconds. Also, in ambulance care the route choice depends on the locations of other available ambulances, which is not the case for hazmat transports.

In this paper we use the DMEXCLP model for the calculation of the OD-pairs, and base our route choice on the MEXCLP model (see [10] for more details). An outline of DMEXCLP follows (an extensive description can be found in [14]). The basic idea of DMEXCLP is to take multiple coverage into account, and consequently, sending an ambulance to the base location where its marginal contribution is the highest. To this end, a constant busy fraction q is introduced, that is the average fraction of the time that an ambulance attends an incident. The marginal contribution to the coverage of the k^{th} ambulance on demand point i (denoted by k_i) is given by $C_i = d_i(1-q)q^{k_i-1}$. Summation over all demand points yields the total contribution of an ambulance movement.

3 Model

The DMEXCLP-model from the previous section, or any other DAM-model, provides the origin O and destination D for ambulance a to send. The goal of our model is to optimize *the route* for ambulance a, while taking transitory coverage into account.

The ambulance region is discretized in a set of demand points $\mathcal{I}$ (where demand point i has demand demand d_i) and base locations $\mathcal{J}$. The set of *waypoints*, denoted by $\mathcal{Y}$, is the set of the locations of all road intersections and all locations where a road changes direction—a curve in the road can be modeled by placing multiple waypoints. Take $\mathcal{L} := \mathcal{Y} \cup \mathcal{I} \cup \mathcal{J}$ as the set containing the waypoints, the demand points, and with the base locations. Denote the minimum travel time between two points $\ell_1, \ell_2 \in \mathcal{L}$ by t_{ℓ_1,ℓ_2}. The road network is modeled as a directed complete graph with nodes $w \in \mathcal{W}$ and the minimum travel times as the edge weights. We assume that the grid of demand points is dense enough in relation to the travel speeds.

A *route* is modeled by a finite sequence $r = (\ell_0, \ell_1, \ldots, \ell_{|r|-1})$ that has taken all its elements from $\mathcal{L}$. A set of routes is referred to as a *routeset*, and is denoted by $\mathcal{R}$.

The proposed dynamic routing method consists of two phases. First, we generate a routeset that contains a fixed number of alternative routes from O to D (see Section 3.1). Subsequent, we evaluate each of the routes in this routeset individually (see Section 3.2). The ambulance is sent over the highest valued route alternative.

3.1 Generating routesets

In this section we show how to calculate the routeset $\mathcal{R}$ for a given OD-pair. In this context, there is a trade-off: on the one hand we need enough routes in the routeset to make a good choice, but on the other hand lead too many routes to unacceptably long calculation times during the evaluation. Therefore, the limited number of routes must be sufficiently different to cover the entire area between O and D.

From practice we get two constraints on a route:

1. It should be easy to explain the route in words over the telephone to an ambulance driver, e.g., a route alternative may not meander through residential areas.

2. EMTs do not accept 'large detours' to reach their destination.

In any case, we want to evaluate the fastest route between O and D as a route alternative. Hence, this is the first route that we include in the routeset.

3.1.1 Decision points

We do not follow the approach from most models in the literature that changes the arc properties for the entire road network in the calculation of a route alternative, as this is too calculation intensive. Instead, we introduce *decision points* on the road network that help us generate more routes, denoted by $v \in \mathcal{Z}$ ($\mathcal{Z} \subseteq \mathcal{Y}$). Decision points are waypoints that lay on the road network where a main road splits, and are used as *via-points*. That is, a route alternative is a combination of the fastest route from the origin to the decision point, and the fastest route from the decision point to the destination. A method that calculates the decision points follows in Section 4. We allow route alternatives to be outside the ambulance region for a limited time. The (partial) travel time matrix with entries t_{wi} is precomputed ($w \in \mathcal{W}, i \in \mathcal{I}$).

This satisfies the first constraint: decision points are easy to explain. As an illustration Figure 1 shows the decision points for ambulance region Gooi & Vechtstreek.

The second constraint motivates to take a subset of the decision points between O and D, and only to consider these decision points as a via-point:

$$\mathcal{Z}_{O,D} := \{z \in \mathcal{Z} : t_{O,z} \leq t_{O,D} \text{ and } t_{z,D} \leq t_{O,D}\}. \tag{1}$$

Figure 2 shows the resulting decision points between Hilversum and Weesp.

Figure 1 The decision points for ambulance region Gooi & Vechtstreek are indicated by black crosses.

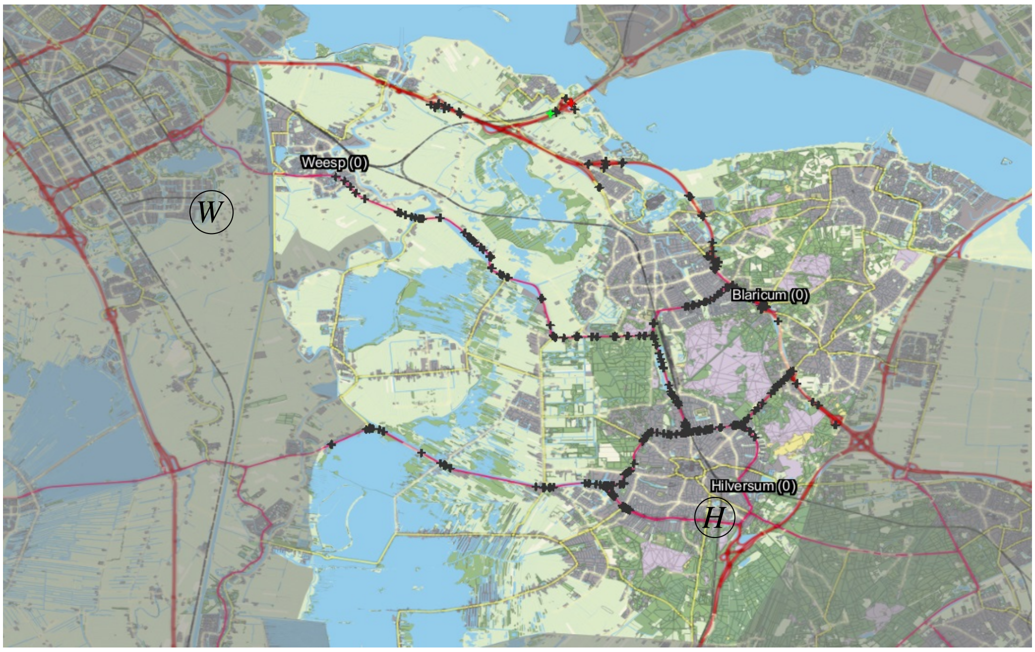


Figure 2 The decision points between Hilversum (H) and Weesp (W).

3.1.2 Snapping

We apply the so-called *snapping* to a route alternative. This procedure is designed to prevent going back and forth over the same path when we create the shortest path through the decision point by removing paths that we travel over in opposite directions. We limit snapping to S minutes: if the original via-point is not reachable by its resulting route alternative, we communicate the point of snapping in its place.

Figure 3 illustrates snapping. While generating route alternatives between Hilversum (H) and Weesp (W), a decision point next to Muiderberg (M) is taken as a pivot. Figure 3a shows the route alternative without snapping. This route alternative is not acceptable for EMTs because it (locally) is a large detour: the same road is traveled twice while entering and leaving Muiderberg. In Figure 3b we applied snapping, and we do not go into Muiderberg, but we still cover the area because the ambulance drives past it.

Algorithm 1 describes the snapping process. We assume two waypoints on opposite sides of the

(a) Route without snapping. **(b)** Route with snapping.

Figure 3 Illustration of snapping on a route from Hilversum (H) to Weesp (W) through Muiderberg (M).

Algorithm 1 The snapping algorithm

$r_1 \leftarrow$ routepart of r from (not incl) O to (not incl) p
$r_2 \leftarrow$ routepart of r's inversed route from (not incl) D to (not incl) p
for all $\ell_1 \in r_1$ **do** $\triangleright$ visit each waypoint ℓ_1 on r_1 in order
    for all $\ell_2 \in r_2$ **do** $\triangleright$ visit each waypoint on the inversed route in order
        $d_{snap} \leftarrow \text{dist}(\ell_1, \ell_2)$. $\triangleright$ Euclidean distance in meters
        if $d_{snap} \leq \Delta$ **then** $\triangleright$ points are within the snapping distance threshold
            if $t_{z_1 p} \leq S$ and $t_{p z_2} \leq S$ **then** $\triangleright$ pivot p remains covered
                return append(route(O, ℓ_1), route(ℓ_1, ℓ_2), route(ℓ_2, D))
            end if $\triangleright$ the route function gives the fastest route
        end if
    end for
end for

road if their difference is at most Δ meter.

3.1.3 Generation algorithm

From hereon we explain the workings of our *routeset generation algorithm*.

For each decision point in $\mathcal{Z}_{O,D}$ we keep track whether there is already a route that visits the decision point: in the case that any route from the routeset visited a decision point, this decision point is marked as *visited*.

After adding the fastest route and marking all the decision points along its way as visited, we take an unvisited decision point as the pivot p. The next route alternative we consider to add to the routeset is the fastest route from O via p to D. We mark all decision points on this new route alternative as visited. For each route alternative r that is already in the routeset, we calculate the *fraction of overlap* with the new route. That is, the number of decision points that is both in the new route and r divided by the total number of decision points that the new route has. If the fraction of overlap is below a given threshold δ for all routes in the routeset, we consider the new route to be sufficiently unique. Only then, we add this route to our routeset. We repeat this process until all decision points in $\mathcal{Z}_{O,D}$ are marked visited.

We use a method to speed up routeset generation, which prevents picking multiple decision points on a road facing outward. Figure 2 illustrates the motivation for the so-called *counter-clockwise pivot picking strategy*. If one first takes the route through the green decision point, one will not reach the red decision point, and thus we have to look at a route through the red decision point at a later time as well, since it is still unvisited. This route will be very similar to the route through the green decision point, and we want to avoid similar routes. Thus, by picking the more outward decision points first, we mark more decision points and prevent looking at too many similar routes.

The counter-clockwise pivot picking strategy is as follows. We first choose the unvisited decision point with the lowest y-coordinate as our pivot. As the next pivot we take the unvisited decision point with the highest x-coordinate. Next we take the highest y-coordinate, and at last, the lowest x-coordinate. Then we look again at the unvisited decision point with the lowest y-coordinate, and we repeat this procedure until all decision points are visited. Hence, we

Algorithm 2 Routeset generation

calculate $\mathcal{Z}_{O,D}$ ▷ use the definition from Equation (1)
mark all decision points in $\mathcal{Z}_{O,D}$ as not visited
calculate route(O, D) ▷ fastest route from O to D
mark all decision points on Route(O, D) as visited
put route(O, D) in the routeset $\mathcal{R}$
repeat
    $p \leftarrow$ first unvisited decision point according to the pivot picking strategy
    $PivotRoute \leftarrow$ append(route(O, p), route(p, D))
    mark all decision points on $PivotRoute$ as visited
    $PivotRoute \leftarrow$ snapping($PivotRoute$) ▷ using Algorithm 1
    $FractionOfOverlap \leftarrow \max_{r \in \mathcal{R}}\{|PivotRoute \cap r|/|PivotRoute|\}$
    if $FractionOfOverlap \leq \delta$ **then**
        put $PivotRoute$ in routeset $\mathcal{R}$
    end if
until every demand point in $\mathcal{Z}_{O,D}$ is marked as visited
return $\mathcal{R}$

mark the decision points in a counter-clockwise fashion, starting from the outside and working towards the inside of the ambulance region.

Lower values of δ result in smaller routesets. Figures 4a, 4b and 4c show multiple routesets that are generated between Hilversum and Weesp for different values of δ. The thickness of the black line indicates the number of route alternatives that use the road segment.

Algorithm 2 concludes the routeset generation algorithm.

3.2 Route choice model

The previous section provided a routeset $\mathcal{R}$. This section shows how a coverage value can be assigned to each route alternative. The route alternative with the largest value is advised to the ambulance.

3.2.1 Outline

Recall that incidents are aggregated to demand points $i \in \mathcal{I}$, and the route alternative visits various waypoints in order, denoted by the sequence $r = (O, \ell_1, \ell_2, \ldots, D)$. These transitional waypoints are not limited to the origin, the decision points and the destination. All changes of travel speeds occur at a waypoint. This is, the speed between two neighboring waypoints is assumed to be constant.

The coverage value of route r can be approximated by adding two terms: (1) the transitional coverage while being on route,* and (2) the coverage when being at the destination.

*For the first term, that is the contribution to the transitional coverage while being on route, we integrate over the travel time θ during the route. We split the integral over the entire route into $\ell - 1$ integrals, one for each road segment. Next, for each point $x(\tau)$ on the route r we calculate the marginal contribution to each demand point i as defined by function $f(\cdot)$, where $t_{x(\theta),i}$ denotes the travel time from the position on the route $x(\theta)$ at time θ to i.

The resulting coverage value of route r is given by the expression:

$$\Xi_r = \sum_{\substack{\ell \in r \\ \ell \neq O}} \sum_{i \in \mathcal{I}} \int_{\tilde{t}_{\ell-1}}^{\tilde{t}_\ell} f(t_{x(\theta),i}) e^{-\gamma\theta} d\theta + \sum_{i \in \mathcal{I}} \int_{\tilde{t}_D}^{\infty} f(t_{D,i}) e^{-\gamma\theta} d\theta. \tag{2}$$

Here, the preceding waypoint to ℓ is denoted by $\ell - 1$. The travel time from O to v is $\tilde{t}_v$. Variable $x(\tau)$ is the position of the ambulance at time τ, that is a linear interpolation in time and space between ℓ and $\ell - 1$. Discount parameter γ models the fact that uncertainty increases as time goes on. Value function $f(\tau)$ gives the marginal contribution of the relocating ambulance to the coverage value as a function of the driving time τ. The integral provides a fair comparison between various routes, because the coverage at the destination weighs heavier if the ambulance arrives sooner.

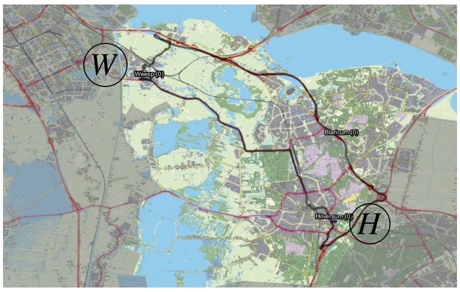
3.2.2 The coverage value for the maximum expected coverage location problem

Recall that the marginal coverage function is denoted by $f(\tau)$ for travel time τ. Function f depends on the location of the other ambulances. For MEXCLP we can use the known result

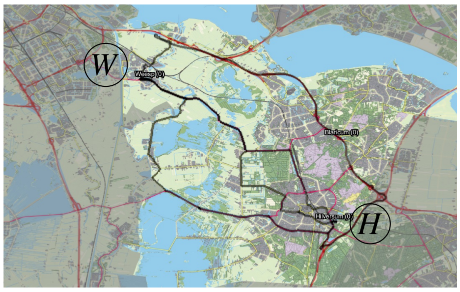
$$f(\tau) = \mathbb{1}_{\{\tau \leq R\}} d_i (1-q) q^{k_i - 1}, \tag{3}$$

where $\mathbb{1}_E$ denotes the indicator function on the event E, q is the average ambulance busy fraction, d_i is the demand at i, and the relocating ambulance is the k_i-th ambulance that can reach i within time threshold R.

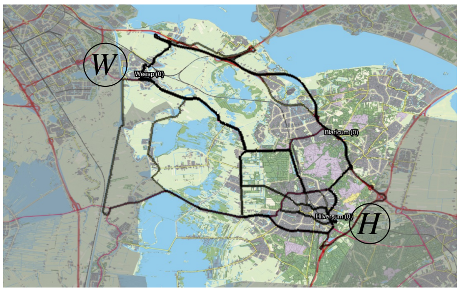
Inspired by literature [14, 5], we fix the locations of all other ambulances $\mathcal{A}^-$ at their current position for the ease of calculations. That is, if they are on a main road we teleport them



(a) $\delta = 0.1$, two routes were added.



(b) $\delta = 0.5$, four routes were added.



(c) $\delta = 0.9$, sixteen routes were added.

Figure 4 Routesets between Hilversum (H) and Weesp (W) with different values for δ.

to the next decision point, and when they are in a residential area we teleport them to the closest demand point. Using the preprocessed travel time matrix, we can rapidly calculate $k_i = 1 + \sum_{a \in \mathcal{A}^-} \mathbb{1}_{\{t_{a,i} \leq R\}}$ the number of ambulances that can reach i within time R, in the case that the relocating ambulance can be on-time at i ($i \in \mathcal{I}$). We compute the contribution if the relocating ambulance is at most R time units away from i, which is independent of the path chosen:

$$C_i = d_i(1-q)q^{k_i-1}.$$

From hereon, we evaluate each route alternative $r \in \mathcal{R}$ in the routeset. For waypoints $\ell \in r$ on the main road network we calculate the coverage from the next decision point the ambulance visits, where we make a correction by subtracting the driving time towards this decision point from the value function's argument. For all other waypoints, which are usually waypoints in residential areas, we calculate the coverage as if the ambulance is at the closest demand point. This is a good approximation because the demand point aggregation is assumed sufficiently dense. Note that by this method the coverage calculation from each waypoint is always calculated from an element in $\mathcal{W}$.

For each pair of demand point i and the route segment preceding ℓ, we compute by linear interpolation the number of seconds that the ambulance is within R time units from i, while driving on this route segment. We assume that the travel speed does not change between two waypoints, which allows us to perform a linear interpolation inside the route segment. Hence, when substitution Equation (3) in the first term of Equation (2), the summation over all road segments of the route r gives the following *contribution* $\xi_{r,i}$ to i:

$$\begin{aligned}
\xi_{r,i} &= \sum_{\substack{\ell \in r \\ \ell \neq O}} \int_{\tilde{t}_{\ell-1}}^{\tilde{t}_\ell} f(t_{x(\theta),i}) e^{-\gamma\theta} d\theta = \sum_{\substack{\ell \in r \\ \ell \neq O}} \int_{\tilde{t}_{\ell-1}}^{\tilde{t}_\ell} \mathbb{1}_{\{t_{x(\theta),i} \leq R\}} C_i e^{-\gamma\theta} d\theta \\
&= C_i \sum_{\substack{\ell \in r \\ \ell \neq O}} \int_{\tilde{t}_{\ell-1}}^{\tilde{t}_\ell} \mathbb{1}_{\{t_{x(\theta),i} \leq R\}} e^{-\gamma\theta} d\theta = C_i \sum_{\substack{\ell \in r \\ \ell \neq O}} \int_{lb(\ell-1,\ell,i)}^{\tilde{t}_\ell} e^{-\gamma\theta} d\theta \\
&= \frac{C_i}{\gamma} \sum_{\substack{\ell \in r \\ \ell \neq O}} \left(e^{-\gamma\, lb(\ell-1,\ell,i)} - e^{-\gamma \tilde{t}_\ell} \right).
\end{aligned}$$

Here the lower bound $lb(\ell-1,\ell,i)$ is the time on the road segment between $\ell-1$ and ℓ when the ambulance driving on this segment becomes within the R time units driving of i. The time measure starts at the route's origin. If i cannot be reached in R time units from the entire segments, we say $lb(\ell-1,\ell,i) = t_\ell$, which results in a zero contribution of this segment. Recall that the travel speed does not change on a road segment that lays between two waypoints. We get the following expression for lb:

$$lb(\ell-1,\ell,i) = \begin{cases} \tilde{t}_\ell & R < t_{\ell,i}, \\ \tilde{t}_{\ell-1} & t_{\ell-1,i} \leq R \text{ and } t_{\ell,i} \leq R, \\ \tilde{t}_{\ell-1} + \frac{R - t_{\ell-1,i}}{t_{\ell,i} - t_{\ell-1,i}} (\tilde{t}_\ell - \tilde{t}_{\ell-1}) & R < t_{\ell-1,i} \text{ and } t_{\ell,i} \leq R. \end{cases}$$

Here, we use the assumption that the ambulance has a constant speed between two waypoints. For the moment of arrival at the destination we get:

$$\int_{\tilde{t}_D}^{\infty} f(t_{D,i}) e^{-\gamma\theta} d\theta = \mathbb{1}_{\{t_{D,i} \leq R\}} \frac{C_i}{\gamma} e^{-\gamma \tilde{t}_D}.$$

This gives us the simplified *path contribution* for the MEXCLP coverage function:

$$\begin{aligned}\Xi_r &= \sum_{\substack{\ell \in r\\ \ell \neq O}} \sum_{i\in\mathcal{I}} \left(\frac{C_i}{\gamma}\left(e^{-\gamma\, lb(\ell-1,\ell,i)} - e^{-\gamma \tilde{t}_\ell}\right)\right) + \sum_{i\in\mathcal{I}} \mathbb{1}_{\{t_{D,i}\leq R\}} \frac{C_i}{\gamma} e^{-\gamma \tilde{t}_D}\\ &= \frac{1}{\gamma}\sum_{i\in\mathcal{I}} C_i \left(\left(\sum_{\substack{\ell \in r\\ \ell \neq O}} \left(e^{-\gamma\, lb(\ell-1,\ell,i)} - e^{-\gamma \tilde{t}_\ell}\right)\right) + \mathbb{1}_{\{t_{D,i}\leq R\}} e^{-\gamma \tilde{t}_D}\right).\end{aligned}$$

4 Results

In this section we show simulation results for the ambulance regions Gooi & Vechtstreek, Amsterdam-Waterland, and Utrecht. We compare the *fastest route* to the *best route generated by the MEXCLP dynamic routing policy*. As key performance indicators we use the fraction of late arrivals and the mean response time. In both policies we use DMEXCLP for the calculation of the OD-pair.

4.0.1 Setup

We use a so-called *trace-driven simulation* strategy for the months September and October 2015. In these months there are no major holidays. In a trace-driven simulation we simulate incidents at exact same time and place as they occurred in reality. The only difference is the way we relocate the ambulances.

A hexagonal equidistant demand point grid is used, such that there is a 1 kilometer distance between the grid points. Demand patterns are calculated from one year historical data. We use $\delta = 0.5$, $q = 0.3$ and $\gamma = 1/(\lambda M)$ the expected inter-arrival time for a particular ambulance. Here, M is the number of ambulances in the region. The Open Source Routing Machine (OSRM) [16] is used for navigation—this software is constructed such that the speed does not change between two waypoints.

From OpenStreetMap (OSM) [17], we create the set of decision points on the major road network $\mathcal{Z}$. These decisions points are identified in OSM as the nodes that lay on a way with highway tag highway, trunk or primary (and their resp. links) that either intersect with another road type or are included in at least three ways. In the Netherlands this results in 49,887 decision points. The choice of including primary but not secondary roads is made because it is generally not hard to turn roads of type secondary or lower. We assume that ambulances pass the following decision point when being dispatched to an incident. For each region, we truncate the set $\mathcal{Z}$ to demand points that are either in the region, or are less than 30 km away from any point in the region. The choice is based such that all demand points of the region are unreachable in 12 minutes driving when not exceeding 150 km/h. This significantly reduces the set $\mathcal{Z}$.

Table 1 shows the simulation results for the *fastest route* and *dynamic routing* policies. In the remainder of this section we consider each of the three ambulance regions separately.

	Fastest route		Dynamic routing	
	Fraction of late arrivals (%)	Mean response time (min:sec)	Fraction of late arrivals (%)	Mean response time (min:sec)
Gooi & Vechtstreek	15.03	11:23	15.87	11:28
Amsterdam-Waterland	0.69	7:50	0.67	7:38
Utrecht	7.03	9:32	7.15	9:28

Table 1 Simulation results for the three regions.

4.1 Gooi & Vechtstreek

The EMS region Gooi & Vechtstreek is the smallest in the Netherlands, both in size and number of calls. It is a rural area with a population just over 250,000. Most people live in the north and the east of the region. The south-west mainly consists of lakes and forests. Yearly, there are over 18,000 incidents. There are three base locations, situated in Hilversum, Blaricum and Weesp. Figure 5 shows a map of the region with its base locations and call volume distribution.

Table 1 shows that there is a small increase in the mean response time. To get a better understanding of where and when we get the most improvement, we look at Figure 6. Nodes are colored green if the dynamic routing method outperforms the fastest route policy, that is where dynamic routing has less late arrivals. A red node indicates that dynamic routing method performs worse. Deeper analysis gave the insight that the high percentage of late arrivals is due to a shortage of ambulances in the evening, especially in the weekends. At that time there are barely enough ambulances available to handle the incoming calls.

Figures 6 and 8 show that the most improvement is in Wijdemeren, the municipality in the south-west corner of the region. This comes at the cost of more late arrivals in the municipality Gooise Meren, which is located in the north. Dynamic routing thus moves the location of late arrivals to more rural areas, since we now take the highway through Gooise Meren less and instead drive through Wijdemeren.

Because the region is understaffed in the evening, we are interested if our method performs better or worse in the evening compared to the rest of the day. Figure 7 shows the late arrivals in the evening compared to the rest of the day. Here the evening is from 16:00 to midnight.

We observe that dynamic routing performs significantly worse in the evening compared to the rest of the day. Sometimes there is a shortage in the number of available ambulances during the evening, which might be the reason that dynamic routing performs worse. The results might improve if we make the discount parameter γ and the busy fraction q time dependent.

If we focus on the per-municipality statistics, we see a shift in the number of late arrivals. Since there is a relatively low call volume in Wijdemeren, any extra on-time arrival results in a larger relative improvement compared to the

more densely populated areas that have a high call volume. Figure 8 shows this relative improvement.

There is a relative improvement in Wijdemeren and Huizen, while dynamic routing performs worse in Weesp. Note that both Wijdemeren and Huizen do not have a base location. Wijdemeren normally has the lowest percentage on time arrivals. Dynamic routing redistributes the late arrivals so the percentage late arrivals of each municipality gets closer together.

4.2 Amsterdam-Waterland

Amsterdam-Waterland has a higher call volume than any other ambulance region in the Netherlands, as it counts 121,000 incidents a year. Fraction 68% of its 1.30 million inhabitants live in the city Amsterdam. This region is densely populated compared to Gooi & Vechtstreek. Figure 9a shows the region, its base locations and distribution of demand.

Figure 9b shows the difference in the number of late arrivals for both policies, where a green node indicates more on-time arrivals for the dynamic routing policy.

Recall that Table 1 shows that the late arrivals in the ambulance region stay about the same, but the mean response time decreases when we use dynamic routing. This is mainly because dynamic routing has the largest improvement in Amsterdam Zuid-Oost (south-east Amsterdam), shown in Figure 9b. This comes at the expense of the semi-rural areas outside of the city Amsterdam that get more late arrivals, especially at Volendam. Observe that Amsterdam Zuid-Oost does not have a base location for ambulances. Thus dynamic routing sends an ambulance over Amsterdam Zuid-Oost to increase coverage over that part of the region. Nowadays, there is a base located in Amsterdam Zuid-Oost.

This illustrates that dynamic routing can be used to cover an area where one would want a base location, and might even be used to search for appropriate base locations.

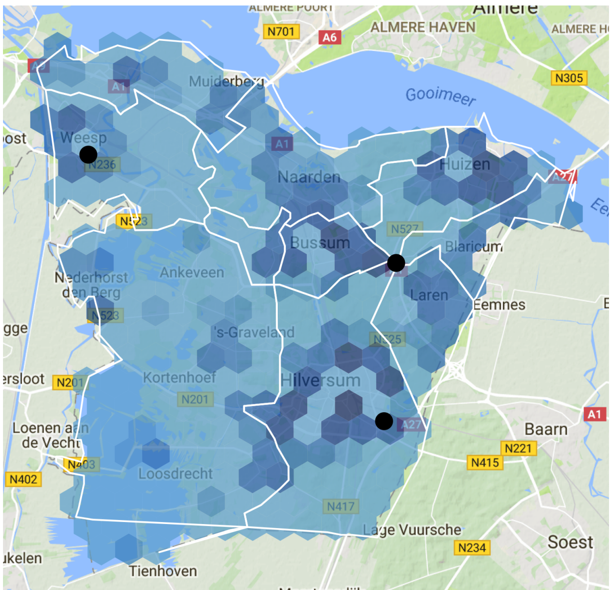


Figure 5 The blue area is the EMS region Gooi & Vechtstreek. The three base locations are indicated by black dots. Darker shades of blue correspond to a higher call volume.

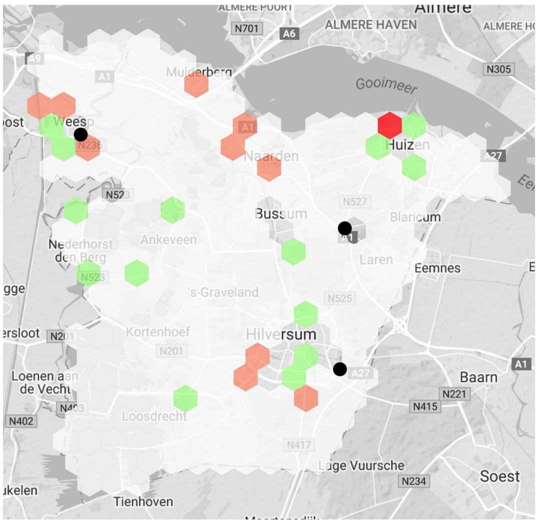
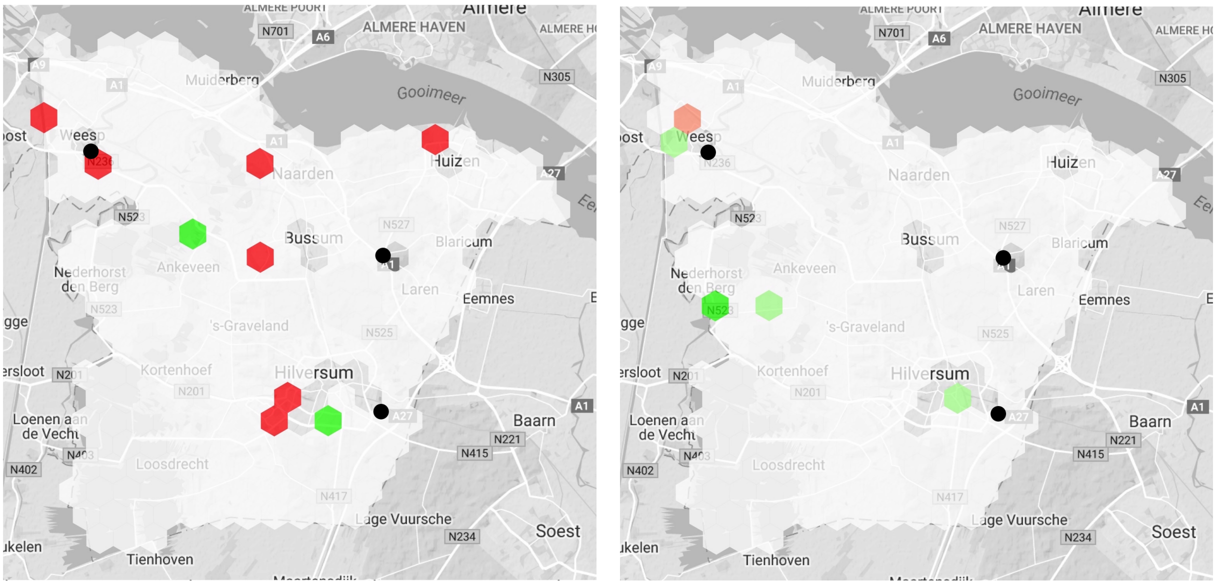


Figure 6 Comparison of the number of late arrivals in Gooi & Vechtstreek. Green areas show improvement with dynamic routing, in contrast to reds. The black dots indicate the base locations.

4.3 Utrecht

Utrecht is a densely populated area with approximately 1.27 million inhabitants. It is amongst the largest EMS regions in the Netherlands. The ambulance provider handles over 90,000 incidents each year. The region and its base locations are shown in Figure 10. We compare both methods similar to the analysis for Gooi & Vechtstreek and Amsterdam-Waterland.

Recall that Table 1 shows a slight increase in the number of late arrivals for the dynamic routing policy. In Figure 11 we see that the most decrease happens in the cities Amersfoort and



(a) Evening.

(b) Rest of the day.

Figure 7 Comparison of the number of late arrivals for different times of the day for workdays. Green areas are improved with dynamic routing, in contrast to reds.

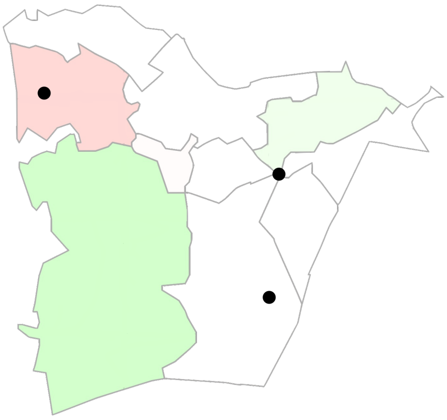
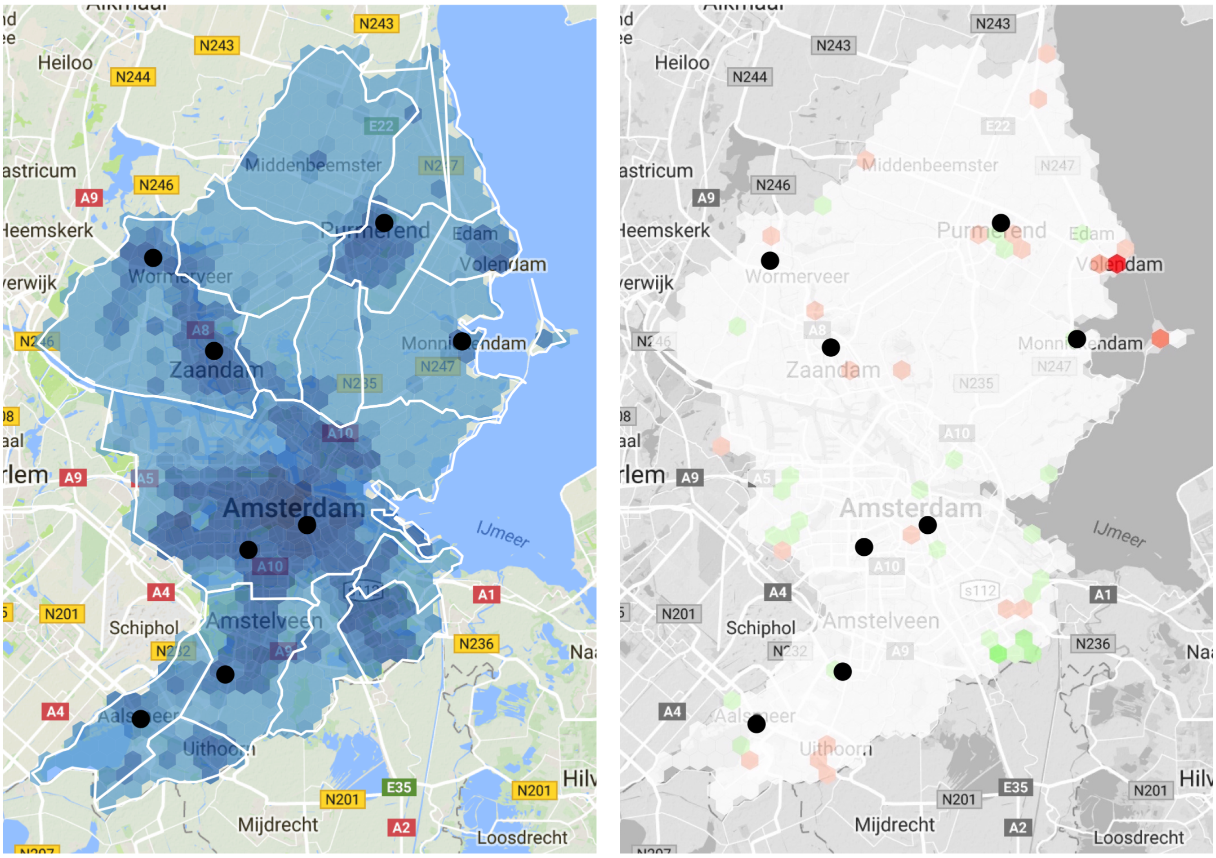


Figure 8 The relative improvement in late arrivals for Gooi & Vechtstreek. Green municipalities show improvement with dynamic routing, in contrast to reds. The black dots indicate the base locations.

Veenendaal. There is a small decrease in the mean response time as well. This can be because the ambulances respond quicker to incidents farther away from base locations when we use



(a) Darker shades of blue correspond to a higher call volume.

(b) The number of late arrivals. Green areas show improvement with dynamic routing, in contrast to reds.

Figure 9 Comparisons for the EMS region Amsterdam-Waterland. The base locations are indicated by black dots.

dynamic routing. The mean response time for this ambulance region is slightly faster.

The most improvement is gained in semi-rural areas with no base location. Especially in Lopik in the south-west and Eemnes in the north-east we see a large increase of on-time arrivals. This is because dynamic routing relocates ambulances through these regions. However, there is a lower performance in the other corners of Utrecht. In both in the north-west and the south-east dynamic routing is outperformed by the fastest route policy. Both these regions have base locations, as opposed to Lopik and Eemnes. Thus because of dynamic routing, the ambulances arrive later at the base locations in these corners of the region, which results in more late arrivals. Hence, we see a redistribution of the late arrivals over the region, where the areas with a lower percentage on time arrivals improve.

Since we have the most improvement in more thinly populated areas, we are interested in the relative improvement of the region. Figure 12 shows the relative improvement for each municipality in Utrecht.

5 Conclusion

Classically ambulances are relocated to a base location using the fastest route. In this chapter we proposed a so-called dynamic routing policy that looks for the best relocation route instead. Simulation results increase the fairness of the ambulance region, while keeping the fraction of late arrivals over the entire ambulance region stable.

Simulations provide interesting additional insights.

First, in the ambulance regions Gooi & Vechtstreek and Utrecht dynamic routing gives more

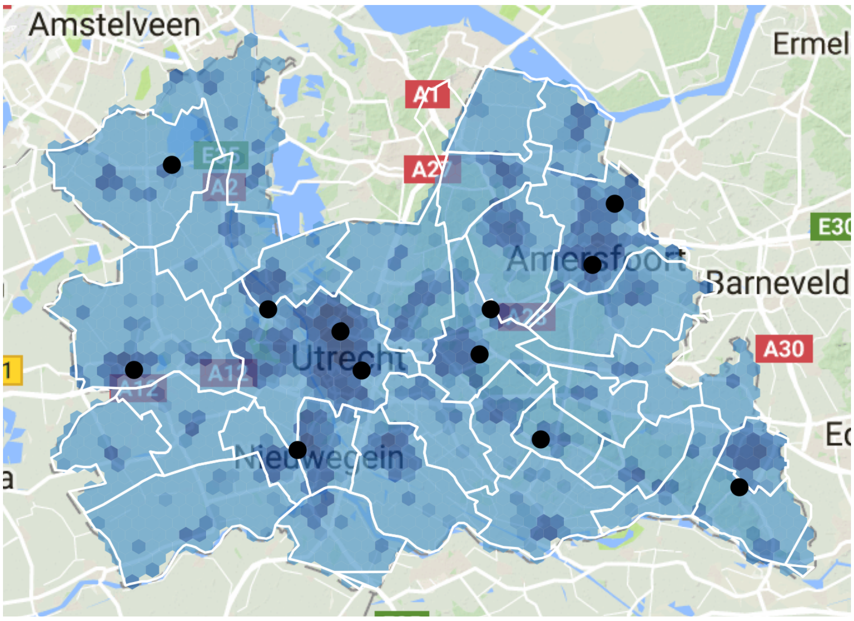


Figure 10 The blue area is the EMS region Utrecht. The base locations are indicated by black dots. Darker shades of blue correspond to a higher call volume.

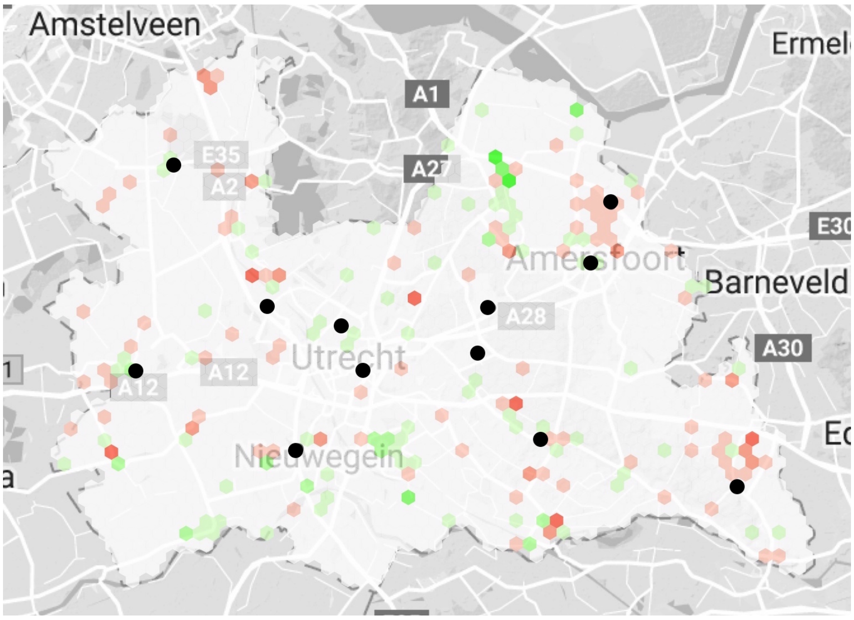


Figure 11 Comparison of the number of late arrivals in Utrecht. Green areas show improvement with dynamic routing, in contrast to reds. The base locations are indicated by black dots.

on-time arrivals in rural areas, at a relatively small cost in the larger cities. The large relative improvement in the urban areas and areas with no base locations shows that dynamic routing ensures a more even distribution of the ambulances.

Second, Amsterdam-Waterland shows that dynamic routing can be used to compensate for sub-optimal locations of ambulance bases. Dynamic routing contributes the most in Amsterdam Zuid-Oost in regards to the number of late ambulance arrivals. This can be explained by a densely populated area without any base location. We note, however, that at the time of writing

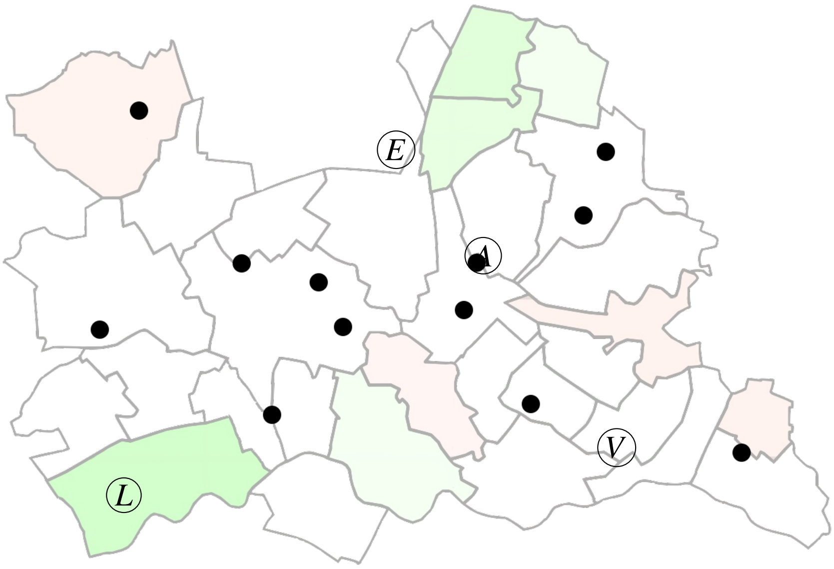


Figure 12 Comparison of the number of late arrivals in Utrecht. Green municipalities show improvement with dynamic routing, in contrast to reds. The base locations are indicated by black dots. Marked are Lopik (L), Eemnes (E), Amersfoort (A) and Veenendaal (V).

a new base location in Zuid-Oost is installed.

Third, an interesting topic for further research is to investigate the effect of the discount parameter γ which reduces the ambulance's contribution on the overall coverage value as time passes. Possibly, there is too much emphasis on the beginning of the route, which is the result of a high choice for the discount parameter. This can lead to ambulances arriving later at their destination, and potentially to unnecessary late arrivals.

Last, it is possible to restrict dynamic routing when certain restrictions are satisfied, i.e., we can only consider dynamic routing during certain times of the day. Further research is needed on the effects of such limitations.

Acknowledgements
Our thanks go to RAVU, Ambulance Amsterdam, and GGD Gooi & Vechtstreek for making the call center records available.

References

[1] Martin van Buuren. "Efficient Planning of Ambulance Services - Theory and Practice, Ch. 3–4". PhD thesis. VU University Amsterdam, May 2018, pp. 49–110. ISBN: 978-94-028-1010-3.

[2] Martin van Buuren, Rob Douwe van der Mei, and Sandjai Bhulai. "Demand-point Constrained EMS Vehicle Allocation Problems for Regions with Both Urban and Rural Areas". In: *To appear in Operations Research for Health Care* (2018). DOI: 10.1016/j.orhc.2017.03.001.

[3] M. Abkowitz, M. Lepofsky, and P. Cheng. "Selecting Criteria for Designating Hazardous Materials Highway Routes". In: *Transportation Research Record* 1333 (July 1992), pp. 30–35.

[4] José Augusto Azevedo et al. "An Algorithm for the Ranking of Shortest Paths". In: *European Journal of Operational Research* 69.1 (Aug. 1993), pp. 97–106. DOI: 10.1016/0377-2217(93)90095-5.

[5] T. C. van Barneveld, S. Bhulai, and R. D. van der Mei. "The Effect of Ambulance Relocations on the Performance of Ambulance Service Providers". In: *European Journal of Operational Research* 252.1 (July 2016), pp. 257–269. DOI: 10.1016/j.ejor.2015.12.022.

[6] M. E. Ben-Akiva et al. "Modelling inter-urban route choice behaviour". In: *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*. Ed. by J. Volmuller and R. Hamerslag. VNU Science Press, Utrecht, July 1984. Chap. 15, pp. 299–330.

[7] Michiel Bliemer and Piet Bovy. "Impact of Route Choice Set on Route Choice Probabilities". In: *Transportation Research Record* 2076 (Dec. 2008), pp. 10–19. DOI: 10.3141/2076-02.

[8] Piet H.L. Bovy and Stella Fiorenzo-Catalano. "Stochastic Route Choice Set Generation: Behavioral and Probabilistic Foundations". In: *Transportmetrica* 3.3 (Jan. 2007), pp. 173–189. DOI: 10.1080/18128600708685672.

[9] Carlos F. Daganzo and Yosef Sheffi. "On Stochastic Models of Traffic Assignment". In: *Transportation Science* 11.3 (Aug. 1977), pp. 253–274. DOI: 10.1287/trsc.11.3.253.

[10] Mark S. Daskin. "A Maximum Expected Covering Location Model: Formulation, Properties and Heuristic Solution". In: *Transportation Science* 17.1 (Feb. 1983), pp. 48–70. DOI: 10.1287/trsc.17.1.48.

[11] Robert B. Dial. "A Probabilistic Multipath Traffic Assignment Model which Obviates Path Enumeration". In: *Transportation Research* 5.2 (June 1971), pp. 83–111. DOI: 10.1016/0041-1647(71)90012-8.

[12] Erhan Erkut and Vedat Verter. "Modeling of Transport Risk for Hazardous Materials". In: *Operations Research* 46.5 (Oct. 1998), pp. 625–642. DOI: 10.1287/opre.46.5.625.

[13] Caroline Fisk. "Some Developments in Equilibrium Traffic Assignment". In: *Transportation Research Part B: Methodological* 14.3 (Sept. 1980), pp. 243–255. DOI: 10.1016/0191-2615(80)90004-1.

[14] C. J. Jagtenberg, S. Bhulai, and R. D. van der Mei. "An Efficient Heuristic for Real-time Ambulance Redeployment". In: *Operations Research for Health Care* 4 (Mar. 2015), pp. 27–35. DOI: 10.1016/j.orhc.2015.01.001.

[15] Otto Anker Nielsen. "A Stochastic Transit Assignment Model Considering Differences in Passengers Utility Functions". In: *Transportation Research Part B: Methodological* 34.5 (June 2000), pp. 377–402. DOI: 10.1016/s0191-2615(99)00029-6.

[16] Open Source Routing Machine. *(Accessed: 2018-01-21)*. http://project-osrm.org.

[17] OpenStreetMap. *(Accessed: 2018-01-21)*. http://www.openstreetmap.org.

[18] Warren B. Powell and Yosef Sheffi. "The Convergence of Equilibrium Algorithms with Predetermined Step Sizes". In: *Transportation Science* 16.1 (Feb. 1982), pp. 45–55. DOI: 10.1287/trsc.16.1.45.

[19] Joseph N. Prashker and Bekhor Shlomo. "Route Choice Models Used in the Stochastic User Equilibrium Problem: A Review". In: *Transport Reviews* (4 July 2004), pp. 437–463. DOI: 10.1080/0144164042000181707.

[20] Carlo G. Prato and Shlomo Bekhor. "Modeling Route Choice Behavior: How Relevant is the Composition of Choice Set?" In: *Transportation Research Record* 2003 (Jan. 2007), pp. 64–73. DOI: 10.3141/2003-09.

[21] Michael Scott Ramming. "Network Knowledge and Route Choice". PhD thesis. MIT, Feb. 2002.

[22] Nadine Rieser-Schüssler, Michael Balmer, and Kay W. Axhausen. "Route choice sets for very high-resolution data". In: *Transportmetrica A: Transport Science* 9.9 (Oct. 2013), pp. 825–845. DOI: 10.1080/18128602.2012.671383.

[23] F. Frank Saccomanno and A. Y.-W. Chan. "Economic Evaluation of Routing Strategies for Hazardous Road Shipments." In: *Transportation Research Record* 1020 (1985), pp. 12–18.

[24] Raj A. Sivakumar, Rajan Batta, and Mark H. Karwan. "A Multiple Route Conditional Risk Model For Transporting Hazardous Materials". In: *Information Systems and Operational Research* 33.1 (Feb. 1995), pp. 20–33. DOI: 10.1080/03155986.1995.11732264.