

## **EVALUATING DYNAMIC DISPATCH STRATEGIES FOR EMERGENCY MEDICAL SERVICES: TIFAR SIMULATION TOOL**

Martin van Buuren

Center for Mathematics and Computer Science  
Science Park 123  
1098 XG Amsterdam, NETHERLANDS

Karen Aardal

Technical University Delft  
Mekelweg 4  
2628 CD Delft, NETHERLANDS

Rob van der Mei

Center for Mathematics and Computer Science  
Science Park 123  
1098 XG Amsterdam, NETHERLANDS

Henk Post

Connexxion  
Laapersveld 75  
1213 VB Hilversum, NETHERLANDS

### **ABSTRACT**

In life-threatening emergency situations, the ability of emergency medical service (EMS) providers to arrive at the emergency scene within a few minutes may make the difference between survival or death. To realize such extremely short response times at affordable cost, efficient planning of EMS systems is crucial. In this article we will discuss the Testing Interface For Ambulance Research (TIFAR) simulation tool that can be used to evaluate the effectiveness of different dispatch strategies. The accuracy of TIFAR is assessed by comparing the TIFAR-based performance indicators against a real EMS system in the Netherlands. The results show that TIFAR performs extremely well.

### **1 Literature Review**

In this section we give a brief overview of the available literature on models for EMS systems.

#### *Deterministic static models*

In the context of EMS, deterministic static models are usually integer programming (IP) formulations for finding a static distribution of EMS vehicles over all potential base locations. The Location Set Covering Model (LSCM) by (Toregas, Swain, ReVelle, and Bergman 1971) is a binary integer programming (BIP) formulation that tries to minimize the amount of bases that contains an EMS vehicle. This must be done in such a way that at most one EMS vehicle will be stationed on a potential base location and each demand point can be reached by an EMS vehicle within a given time  $r$ . LSCM gives a lower bound for the amount of EMS vehicles needed. The Maximal Covering Location Problem (MCLP) by (Church and ReVelle 1974) is a BIP-formulation that tries to find a way to distribute a given number of EMS vehicles over the potential base locations such that as many demand points can be reached by at least one EMS vehicle within the given time standard  $r$ . The major disadvantage of both LSCM and MCLP is that at the moment when an EMS vehicle departs to an incident location some demand points might not get reached within the given time standard, or rather said: these demand points might not be covered any more. The Backup Covering Problems (BCP) in (Hogan and ReVelle 1986) try to solve this by rewarding double covered demand points in the objective function. The Double Standard Model (DSM) in (Gendreau, Laporte, and Semet 1997) can be seen as an extension of BCP. This model maximizes the population covered by (at least) two EMS vehicles, and it is the first model that considers two different maximum allowed patient

waiting times  $r_1 < r_2$ , from now on called response times, for high urgency and lower urgency calls, respectively. In case there are two types of vehicles, advanced live support (ALS) vehicles that may handle all types of calls and basic live support (BLS) vehicles that may only handle lower urgency calls, the Tandem Equipment Allocation Model (TEAM) in (Schilling, Elzinga, Cohon, Church, and ReVelle 1979) gives a way to model the situation. This model maximizes the population covered at least once by each type of vehicle given the number of EMS vehicles of each type and the corresponding response times  $r_1$  and  $r_2$ .

#### *Probabilistic static models*

Probabilistic static models are more realistic in the sense that they take the probability  $q$  into account by which an EMS vehicle is available for dispatch, independent of the status of all other EMS vehicles. The Maximum Expected Covering Location Problem (MEXCLP) in (Daskin 1983) gives a lower bound for the amount of vehicles that is needed, like LSCM does for deterministic static models. In case one wants to assign a given number of EMS vehicles to potential base locations such that a maximal number of demand points can be reached with some predetermined probability  $\beta$  within a given time standard  $r$ , one can use the Maximum Availability Location Problem (MALP) in (ReVelle and Hogan 1989). The objective of Rel-P in (Ball and Lin 1993) is to impose an upper bound  $\beta_i$  on the probability that a call on demand point  $i$  does not receive immediate service. This model also allows an upper bound for the number of EMS vehicles on each base. The Two-Tiered Model (TTM) in (Mandell 1998) can be seen as the probabilistic static version of TEAM.

#### *Dynamic relocation models*

Modern research in EMS deployment is mainly focused on dynamic EMS modeling. Dynamic models help to relocate idle EMS vehicles such that a maximum of calls can be reached within a time threshold. In relation to the previous models, dynamic models are not searching for a static equilibrium but rather contribute to real life relocation systems. The Dynamic Double Standard Model (DDSM') in (Gendreau, Laporte, and Semet 2001) allows relocations with an induced cost of available EMS vehicles between bases, and its objective function maximizes double coverage whilst relocation costs act as a penalty. The Approximate Dynamic Programming (ADP) formulation as described in (Restrepo 2008) and (Maxwell, Restrepo, Henderson, and Topaloglu 2010) captures the random evolution of a system and uses simulation to try to obtain the best dispatch policy such that the total costs are minimized. Both DDSM' and ADP only search for possible relocations at the moment that an EMS vehicle becomes available or unavailable for dispatch to a new call.

For a more thorough discussion on how these static and dynamic models work we refer the reader to (Buuren 2011) and (Brotcorne, Laporte, and Semet 2003).

#### *Simulations models*

Simulations models are extremely powerful, because of their highly flexibility, e.g. with respect to modifications to the model assumptions. Moreover, they typically provide graphical information that is very illustrative, and therefore are highly appreciated by EMS managers. Also, the effect of certain decisions is easy to understand and explain when a simulation is used. Simulations provide a way to see the effect of certain decisions on a real time basis, see (Henderson and Mason 2005). The effect of all models that are mentioned before are tested on real life environments by using some kind of real data such as historical call record data. Two simulation packages are worth mentioning: BartSim and SIREN. BartSim is a simulation package developed in (Henderson and Mason 2005) for St. Johns Ambulance Service in Auckland, New Zealand, to assist during policy making. EMS vehicles have a computer-aided dispatch system that logs all call data such as travel times, treatment time and transfer time. This simulation engine is the first of its kind that uses real data for modeling the calls. This way, data does not have to be recorded manually as in EMS studies before; Henderson and Mason state that during the survey in (Swoveland, Uyeno, Vertinsky,

and Vickson 1973) data was gathered manually for a period of two weeks. They also point out that using a GIS-system is relatively new in EMS planning. Auckland was described by a graph containing 2 200 vertices and 5 000 directed arcs. Incidents are generated on the vertices by a bootstrapping procedure. Some of these vertices are neither intersections nor dead ends. Leaving them out of the graph results in 765 vertices called decision vertices. The travel speed of the EMS vehicles is time dependent in BartSim. During pre-processing, the all-to-all shortest path between these 765 decision vertices is calculated with the Floyd-Warshall algorithm at times 8:00, 12:00 and 17:00, where a heuristic is used to estimate the travel speed at these times. See (Floyd 1962) for more information about the algorithm. Travel times between these three times are calculated by interpolation. The shortest path calculation between two vertices are done by a heuristic that is claimed to have a good level of accuracy. The BartSim simulator works as a discrete-event simulator. SIREN is the successor of BartSim and it simulates EMS movements as well. This software package was used for the Auckland and Melbourne areas. Nowadays the simulation package is integrated in commercial packages. Some information about SIREN can be found in (Henderson and Mason 2005) and (Mason 2005). SIREN uses better base locations and it considers to move bases for improved response times. SIREN includes real data and simulations yield an improvement up to 9% on the previous strategy Melbourne used. SIREN also includes stochastic travel times and non-homogeneous call generation. The simulation package can dispatch more than one vehicle to the same call. SIREN can handle up to 6 000 vertices and 14 000 arcs, and contains arc specific travel times for EMS vehicles that drive with optical and auditory signals.

## **2 EMS in The Netherlands**

In this section we give a brief outline of the EMS system in the Netherlands. The Netherlands is partitioned into 25 EMS regions, each with a single ambulance service provider, called RAV (Dutch: regionale ambulancevoorziening). Emergency medical call centers (EMCCs) handle the incoming medical emergency calls by civilians, medical specialists and other emergency services. It also coordinates EMS vehicle movements within the RAV. For a complete overview of demographical data we refer to (CBS 2009).

### *Urgency levels*

Each incoming call is assigned an urgency level. We distinguish three levels: A1, A2 and B.

- A1 An urgent call with an acute threat to the patients life. Vital functions of the patient are not or rarely present, or cannot be determined through the telephone. The EMS vehicle uses optical and visual signals and tries to get to the patient as soon as possible. Examples: Heart attack, reanimation or serious traffic incidents.
- A2 The patients life is not under direct threat, but there might be serious injuries. The EMS vehicle may use optical and visual signals if the EMS personnel has discussed this with the EMCC, but this only happens on rare occasions. Examples: A broken leg or a general practitioner asks for transportation to a hospital.
- B A call without A1 or A2 urgency in which the patient must be transported within a given predetermined time interval. A typical B call exists of transferring a seriously ill person from one hospital to another, because this hospital is specialized in the patient's condition. When a seriously ill person receives a scheduled transport from an EMS vehicle to his or hers home, it will be classified as a B call as well.

A major difference between calls that are labelled with urgencies A1 and A2 on one hand and calls with urgency B on the other, is that calls with A1 or A2 urgency are not known at forehand, whilst calls with B urgency can be planned in advance.

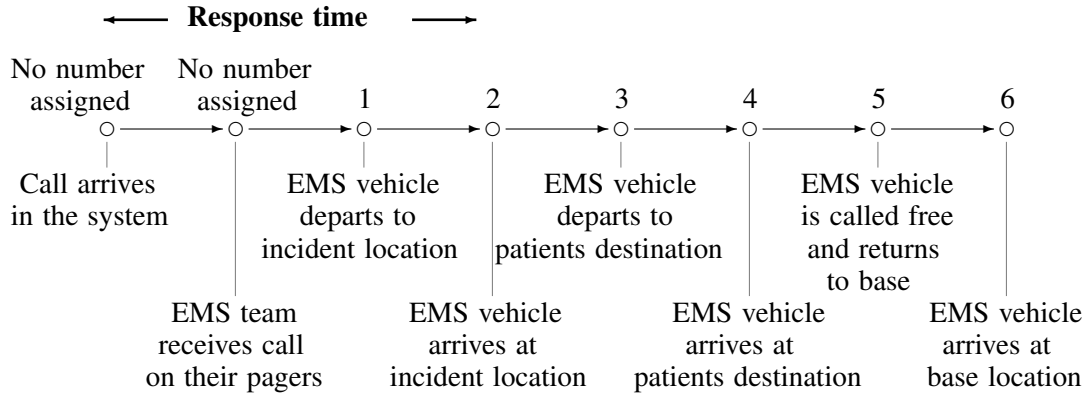


Figure 1: Overview of system statuses.

### Classification of calls

Each call is classified as a *declarable*, *EHGV* or *loss*. A call is classified as declarable if transportation is required. An EHGV-call (Dutch: Eerste hulp, geen vervoer) is a type of call where the EMS team can provide care locally, and the patient does not require transportation to a hospital. This status is determined by EMS personnel at the incident location. For EHGV calls, mobilophone status 5 will immediately follow after status 2. Sometimes an EMS team arrives at the scene and misses the presence of a patient to take care of. This might be the result of a patient who left after the call was made or a prank call to the emergency number. This type of call can be classified as a loss call. Calls that are not EHGV or loss are classified as declarable. In the case a call is declarable a patient will be brought to a hospital.

### Monitoring the status of calls

For each incoming call, status information is time-stamped and logged. See Figure 1 for a systematic overview of all statuses that a call can have. The *response time* is defined as the length of the time interval between the moment that the call center receives the phone call until the moment the EMS vehicle arrives at the incident location, see also Figure 1. The maximum response time of calls with A1 urgency is 15 minutes, and for A2-calls 30 minutes. A key performance indicator is the fraction of calls that meets these maximum response-time thresholds. For B-calls there is not a maximum response time defined, since they are usually planned in advance.

## 3 The Testing Interface For Ambulance Research (TIFAR) simulation package

In this section we give an outline of how TIFAR works. TIFAR is programmed in ansi C++, making it easily transferable to different platforms and easy to perform maintenance. Essentially, the program is based on three connected loops: (1) the main loop, (2) a renewal loop, and (3) an incident generator loop. Also, TIFAR can run in two modes: visual mode or speed simulation mode. The only two ways that these two modes differ are the moments on which they set the next time stamp where the (new) state of the system gets calculated, and the way how they perform output. Visual mode uses a GUI for output and the internal computer clock to determine the next time in which the current system state gets calculated. In this way one can see the EMS movements at real time, or with a speedfactor. Speed simulation mode is a discrete event simulator that holds an ordered list with the end times of all currently ongoing events. For example: At the moment when an EMS vehicle departs, we can calculate the moment when the vehicle arrives. This gives a new end time. An end time enters the list when a call enters the system, when a vehicle departs from a location or when a vehicle arrives at a location. The next time at which the state of the system must be calculated then equals the first one in this ordered list. When a predetermined end time has passed, speed simulation mode terminates the program and displays statistics in the terminal. Speed





Figure 2: Illustration of Amsterdam region with EMS bases and hospitals.

simulation mode is faster than visual mode, because the renewal loop gets only called at necessary time stamps. For each of the two modes there is a separate main loop included in the program.

### Call generation

The incident generator loop generates calls with the following properties:

1. The model time when the incident occurs, called the start time.
2. The location on the map where the incident occurs, called the origin.
3. The treatment time, i.e. the time the EMS personnel must spend at the origin.
4. The transfer time, i.e. the time the EMS personnel must spend at the hospital to transfer the patient into hospital care. Only if the call is not EHG (or loss).
5. The urgency of the call.
6. Whether the call is EHG (or loss) or not. Recall that EHG means that the patient does not require transportation to the hospital.

There are two ways to generate calls: one can choose to give each demand point its own distribution, or one can choose to have one distribution for the moment when a call in the RAV occurs and a separate distribution to determine the origin of a call within the RAV. TIFAR makes use of the latter. Calls are generated by a Poisson process.

For ease of the model, it is assumed that each call is handled by exactly one EMS vehicle; note that this assumption can easily be relaxed. There are a couple of possible choices for generating the call's origin:

- *Using RD-coordinates:* The Netherlands has its own cartesian coordinate system called the rijks-driehoekskoördinaten, where the unit is 1 meter. Every location in the country has an  $x$ - and  $y$ -coordinate. When knowing the border of the RAV in RD-coordinates, one can generate a call on a random position uniformly distributed within the RAV. The main advantage is that an incident can happen on every location within the RAV, even on water. One still has to keep in mind that this location must be mapped upon the road network which might lead to major granularity errors. The main disadvantage is that the population density is not considered in this distribution.
- *Using postal codes:* Every address with a mail box has one postal code assigned. Such a postal code consists of four digits and two letters, for example 1011 AA. There are buildings with the same postal codes: a part of the same street can share a postal code. However, the combination of a postal code and house number forms an unique combination. The four digits are forming the neighborhood, whilst the two letters specify the location within this neighborhood. People involved in route planning in The Netherlands therefore make the distinction between these so called 6PP- and 4PP-postal codes. A rule of thumb states that the cumulative amount of mail for all houses with the same 6PP postal code is the quantity that a post man can hold in his hands. One can map a postal code onto RD-coordinates. The main advantage is that the population density is included. Using the rule of thumb, one can say that the number of people located on each postal code is almost equal. One must keep in mind that streets with only one mail box and not many inhabitants have their own postal code, while on the other hand nursing homes have one postal code and a high potential that an EMS vehicle should be called. The disadvantage is that forests, water and highways do not have postal codes because there are no mail boxes. Still, incidents can happen at these places.
- *Using bootstrapping:* Using EMS data from the past, one knows where an incident has happened, and thus where an incident can happen again. In The Netherlands, all EMS data from 2007 until now is stored. Incidents in the model can be generated using a bootstrap procedure from this data. The main disadvantage is that places where many incidents happen in reality, will be represented very accurately. The disadvantage is that there are a lot of places where no incident has happened before, but where new incidents might occur. Furthermore, new neighborhoods are not included by this way of incident generation. The use of historical EMS data can also involve privacy concerns, though when one can afford to lose some precision, the location can be anonymized by mapping it onto the 4PP postal code.

### *Dispatching policy*

When a call occurs, we have to assign an EMS vehicle to the call. We will now describe how the call handling works. TIFAR has a queue that contains all calls. We assign EMS vehicles to calls *in order of priority*: first we assign vehicles to calls with urgency A1 on a first-come-first-served basis, and if all calls with urgency A1 are served we start assigning vehicles to calls with urgency A2 on the same basis. The EMCC always assigns the nearest (in time) available EMS vehicle to a call. Note that a similar approach has been applied in (Restrepo 2008). Once an EMS vehicle has been assigned to a call, the call will not be handled by another EMS vehicle. Not even when the other EMS vehicle gets available for dispatch while being closer to the origin than the already assigned vehicle. This assumption is based on the fact that in practice, it rarely occurs that an EMS vehicle gets assigned to another call.

When driving to a call with A1 urgency we assume the vehicle has auditory and visual signals, and when driving to a call with A2 urgency we assume that the vehicle drives without them. Sometimes an

EMS vehicle drives with these signals to an A2 call, but since this rarely occurs we have not implemented this in TIFAR. When driving to a hospital, we assume that the EMS vehicle goes with A2 speeds. The speed at which a vehicle moves depends on the type of road it is on (e.g., highway, road in city, road out city) and the (absence of) auditory and visual signals. In reality, an EMS vehicle may drive with signals to an hospital.

When a call is declarable, the patient will be brought to the nearest hospital. We assume that this corresponds well with the real situation, although there are cases in which a specialized hospital should be chosen instead of the nearest one, see (Groen 2009). When an EMS vehicle departs from a hospital, it will head to the nearest base unless a relocation rule decides otherwise.

Relocation policies are confidential, and therefore we use a heuristic. We call an EMS vehicle *involved* with a base if one of the following two conditions hold:

1. The EMS vehicle is waiting at the base until a new call arrives.
2. The EMS vehicle is driving to the base and is available for dispatch or relocation.

If a relocation rule states that an EMS vehicle must be send from base  $j_1$  to base  $j_2$ , TIFAR takes the set of all EMS vehicles involved with  $j_1$ , calculates each of their distances in time to  $j_2$  and sends the nearest one to  $j_2$ . When there are zero EMS vehicles involved with  $j_1$  the EMCC of TIFAR will not relocate an EMS vehicle from  $j_1$ . When relocating, we make a distinction between three regions (namely, North, Center and South, as can be seen in Figure 2).

If an EMS vehicle does not receive a relocation at the moment it gives mobilophone status 5 (see Figure 1), the vehicle will go to the nearest base to wait until it gets assigned to a new call or until a relocation rule at a later time will send it to another base. This strategy has the disadvantage that EMS vehicles get drawn to the centre. Let us illustrate that. In the North, there are two hospitals. When an EMS vehicle delivers a patient to one of the Northern hospitals, it will return to the nearest base which is in the centre. The vehicle is now drawn to the centre. In the south, there is a region where the nearest base is Amstelveen, and the nearest hospital is VU Medical Center (VUmc). After delivering a patient to the VUmc hospital the EMS vehicle will find that a base in the center is nearest, and will go to there to wait until a new call arrives. When a vehicle enters the centre region there are two ways to leave it:

1. By a relocation rule.
2. When an incident occurs in another region and that region has zero available EMS vehicles. Often this leads to calls with A1 urgency that are not served in time.

These observations again stress the importance of good relocation rules. Relocations are necessary to keep available EMS vehicles distributed well over the RAV.

TIFAR makes use of graphical user interface and route planning software. Let us mention the most important services. The map server has as input the RD-coordinates of the bottom left corner, output image dimensions and a scale factor, and returns the corresponding map in JPG format. The position planner has as input a postal code and house number, and returns the corresponding RD-coordinates. This is used to determine the location of the predetermined EMS bases and hospitals, and to display them on the map. The geo-projector maps RD-coordinates to the 'nearest' road. The geo-projector is used to determine the location of an available EMS vehicle that is not on a base. The AB planner returns the shortest route and the corresponding time in seconds between two points in RD-coordinates. As input one gives the start point and end point. This input can be an address, a postal code or a geo-projection. TIFAR uses the latter two options. The EMS vehicles' travel speeds are obtained from historical data. The shortest path is calculated with the bidirectional A\*-algorithm, see (Klunder and Post 2006). The matrix planner has two modes:

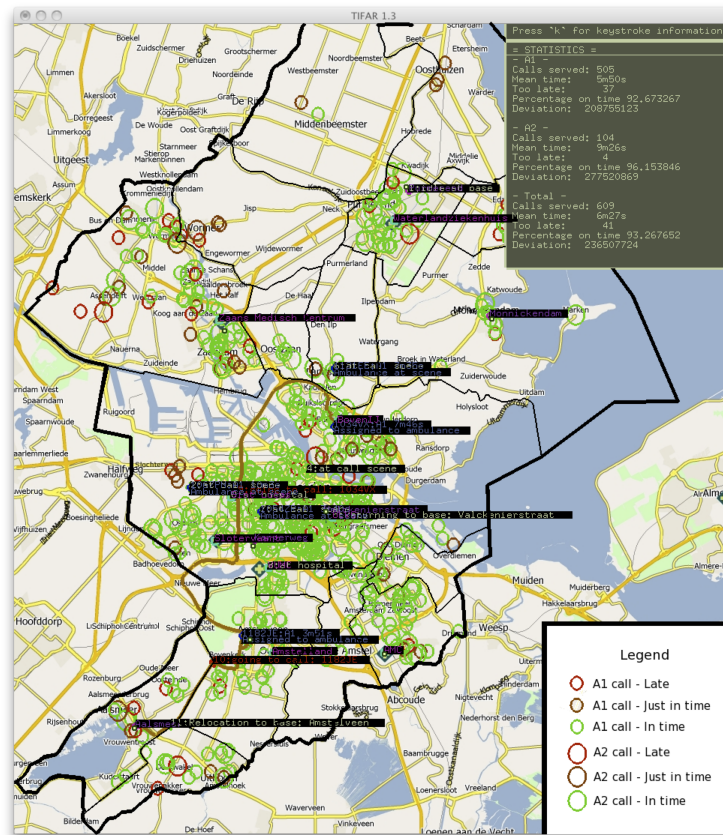


Figure 3: Illustration of the GUI of TIFAR.

one-to-many and many-to-one. The many-to-one determines the distance and travel time from many start locations to only one end location. This is used to determine the nearest available EMS vehicle: from many vehicles to one call's origin. The one-to-many determines the distance and travel time from one location to many locations. This is used to determine the nearest hospital from a call location, or to determine the nearest base from a hospital when an EMS vehicle returns to a base. Input can be both a postal code or geo-projection. Tele Atlas data is used for road network information and geographic information.

## 4 Results

We have run a variety of simulations with realistic scenarios based on the parameters obtained from historical data and from detailed information about the base locations, staffing levels, travel-time models and relocation policies implemented in an operational RAV in the Netherlands. The results show that the simulation-based performance of high-priority calls is extremely close to the actually realized performance by the RAV (the actual numbers are confidential). This shows that TIFAR is able to answer what-if scenarios for different configurations, which provides RAV with a powerful means to enhance the efficiency of their daily operation.

## 5 Conclusion

TIFAR is a decision support tool that is good to predict high urgency calls. Our simulation results for calls with A1 urgency are very close to the actual statistics. TIFAR has several great features: (1) it can hold large amounts of vertices (demand points, bases and hospitals). Over 32 000 vertices are supported, which

leads to very small granularity errors. (2) The graphic user interface can display effects of decisions in real time. (3) It is easy to implement relocation rules. (4) It is easy to vary the amount of EMS vehicles. (5) It gives a clear overview of important statistics.

## 6 Acknowledgements

The authors would like to thank Simon Visser, Maya Ronday, Suzanne van der Leeuw and Rob Bosman for their valuable input.

## REFERENCES

- Ball, M., and F. Lin. 1993. "A reliability model applied to emergency service vehicle location". *Operations Research*:18–36.
- Brotcorne, L., G. Laporte, and F. Semet. 2003. "Ambulance location and relocation models". *European Journal of Operational Research* 147 (3): 451–463.
- Buuren, M. v. 2011, January. "TIFAR modeling package for the evaluation of emergency medical services". Master's thesis, Delft University of Technology.
- CBS 2009. *Demografische kerncijfers per gemeente 2009*. Statistic Netherlands.
- Church, R. L., and C. ReVelle. 1974. "The maximal covering location problem". *Papers in regional science* 32 (1): 101–118.
- Daskin, M. 1983. "A maximum expected covering location model: formulation, properties and heuristic solution". *Transportation Science* 17 (1): 48.
- Floyd, R. 1962. "Algorithm 97: shortest path". *Communications of the ACM* 5 (6): 345.
- Gendreau, M., G. Laporte, and F. Semet. 1997. "Solving an ambulance location model by tabu search". *Location science* 5 (2): 75–88.
- Gendreau, M., G. Laporte, and F. Semet. 2001. "A dynamic model and parallel tabu search heuristic for real-time ambulance relocation". *Parallel computing* 27 (12): 1641–1653.
- Groen, A 2009, December. "Preklinische traumazorg".
- Henderson, S., and A. Mason. 2005. "Ambulance service planning: simulation and data visualisation". *Operations Research and Health Care*:77–102.
- Hogan, K., and C. ReVelle. 1986. "Concepts and applications of backup coverage". *Management Science*:1434–1444.
- Klunder, G., and H. Post. 2006. "The shortest path problem on large-scale real-road networks". *Networks* 48 (4): 182–194.
- Mandell, M. 1998. "Covering models for two-tiered emergency medical services systems". *Location Science* 6 (1-4): 355–368.
- Mason, A. J. 2005, April. "Siren: An Ambulance Simulation System". In *Mathematical Models for Optimizing Transportation Services*: University of Auckland.
- Maxwell, M., M. Restrepo, S. Henderson, and H. Topaloglu. 2010. "Approximate dynamic programming for ambulance redeployment". *INFORMS Journal on Computing* 22 (2): 266–281.
- Restrepo, M. 2008. *Computational methods for static allocation and real-time redeployment of ambulances*. Ph. D. thesis, Cornell University.
- ReVelle, C., and K. Hogan. 1989. "The maximum availability location problem". *Transportation Science* 23 (3): 192.
- Schilling, D., D. Elzinga, J. Cohon, R. Church, and C. ReVelle. 1979. "The TEAM/FLEET models for simultaneous facility and equipment siting". *Transportation Science* 13 (2): 163–175.
- Swoveland, C., D. Uyeno, I. Vertinsky, and R. Vickson. 1973. "Ambulance location: a probabilistic enumeration approach". *Management Science*:686–698.
- Toregas, C., R. Swain, C. ReVelle, and L. Bergman. 1971. "The location of emergency service facilities". *Operations Research*:1363–1373.

## **AUTHOR BIOGRAPHIES**

### **MARTIN VAN BUUREN**

### **KAREN AARDAL**

**ROB VAN DER MEI** is full professor in Operations Research at the VU University Amsterdam, and is heading the research cluster Probability, Networks and Algorithms, consisting of some 70 researchers, at the Centrum Wiskunde & Informatica (CWI). His research is mainly focused on the development and analysis of mathematical models for real-life problems related to capacity planning and performance analysis in many application areas. He has been involved in countless research projects with industrial partners, and is co-author of over 100 papers in the field.

### **HENK POST**