Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

MODELING AMBULANCE SERVICE OF THE AUSTRIAN RED CROSS

Othmar Koch

Vienna University of Technology
Inst. for Appl. Math. and Numer. Anal.
Wiedner Hauptstrasse 8–10
A-1040 Wien, AUSTRIA

Heinz Weigl

ESLA Solutions
Scheibengasse 5
A-1190 Wien, AUSTRIA

ABSTRACT

We discuss a simulation model used in the analysis of the transport logistics of the Austrian Red Cross rescue organization. The emphasis is on the details of modeling the scheduling of ambulance service in the simulation environment ARENA. A heuristic (near-) optimal strategy is employed to coordinate patients' transports, where some parameters with an intuitive interpretation, which are involved in the decision process, have to be suitably chosen. The validity of the model is apparent from the interpretation of the results in terms of the structure of the organization and coordination of services provided.

1 INTRODUCTION

The aim of our analysis of the transport logistics of the Austrian Red Cross described in this paper is to compare the current structure of the organization with an alternative scenario intended for an improvement of the efficiency of ambulance service. To conduct the study we chose to utilize a discrete simulation model, since classical approaches for the optimization of the transport of goods seemed inappropriate for our purpose. The mathematical tools for the analysis of transport problems discussed in Domschke (1989) and Domschke (1990) cannot capture the dynamic situation at full but rely on average (or possibly stochastic) demands and supplies. Our simulation model was implemented in ARENA/SIMAN. The SIMAN simulation engine turned out to be the appropriate tool for our purpose, and the ARENA system provided a comfortable developing environment. However, for reasons explained below we refrained from using any of the ARENA modules, but restricted ourselves to the elements of the SIMAN simulation language. A detailed description of our model and some hints at the implementation are given in §2. In §3 we give the results of our simulation and discuss implications on possible improvements of the coordination of transports.

2 THE ARENA/SIMAN MODEL

The traffic network underlying our model was implemented as a SIMAN *network* consisting of *intersections* (and associated *stations*) and connecting *links*, which also enables the use of *guided transporters* navigating on the graph. The use of *guided transporters* in a *network* implies the possibility to use the graph algorithms integrated in SIMAN to solve shortest path problems (Pegden, Shannon, and Sadowski 1995). However, this is the technical reason why we had to refrain from using any ARENA modules. The use of any of these high-level constructs invokes the activation of a *distance module* (normally used for *free transporters*) which is in conflict with the utilization of a *network* and *guided tranporters*. The network consists of about 300 nodes and 1400 links connecting them. A graphical representation of the graph is given in Figure 1. The two nodes denoted by **H** represent cities with a number of hospitals. Both places are not part of the area we are discussing. The remaining graph is divided into three subareas **Area 1**, **Area 2**, and **Area 3**. Large dots represent places with Red Cross stations.

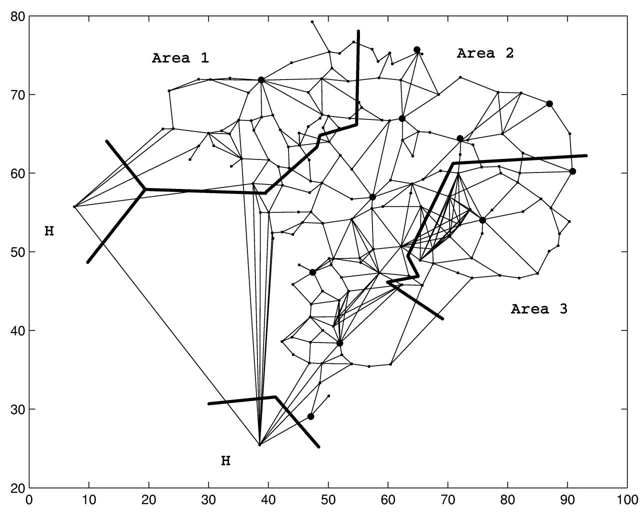

Figure 1: The Traffic Network

For the ambulance service we use three different kinds of transporters with different transport capacities and demands on operating personnel:

- *Notarztwagen* (NAW) transport only one patient, require three persons to operate and are used for emergencies only.
- *Behelfskrankentransportwagen* (BKTW) take up to four patients, which have to be able to walk of their own accord, and only require one driver.
- *Krankentransportwagen* (KTW) can carry up to three patients, one on a stretcher and two on sedan chairs. For the different types of patients, see below. A KTW requires two operators.

The ambulances are stationed in special parking positions representing the locations of the Red Cross stations in the area, cf. Figure 1. For technical reasons, one particular parking station is assigned to each vehicle. The average speed of transporters is assumed to be 60 km/h in general, while for emergency transports 90 km/h is permitted and in cities or villages an average of only 30 km/h is prescribed. The transporters are routed using special *driver entities* which are responsible for the acquisition of operating staff (see below), control of the free capacity, and updating and executing the planned route. To avoid deadlocks, a *relinquish block* is used upon every activation of a transporter to enable the guided transporters to pass each other by uninhibited.

The staff required to operate an ambulance is a *resource* with a capacity that is governed by a schedule. There is a different pool of rescue staff with an individual schedule for every Red Cross station. Assignment of personnel to a transport is *first come first serve*. The mechanism applied when a change in capacity cannot be effected immediately will be discussed later.

The patient data used to drive the simulation is read in from a text file containing the entry time of the patients into the system, that is, the time of the first request for ambulance service, the place (vertex of the network) of entry and the destination of the requested transport. Additionally, the patient type is read in from the file. There are four different categories for this attribute:

1. Emergency patients, who require preferential treatment, see below.
2. Regular patients who are still able to walk of their own accord.
3. Regular patients who have to be carried (and transported) on a stretcher.
4. Regular patients who have to use a sedan chair.

The data was collected for a three month period (January to March 2001) and slightly adjusted to avoid exceptions due to holidays disrupting the duty roster of ambulance personnel. Altogether, a data set of 14,174 patients was used to drive the simulation.

When a patient enters the system, the patient type is determined first. For an emergency patient, an admissible ambulance is assigned to carry through the transport as fast as possible. An ambulance is considered admissible for an emergency if it has free capacity to transport a patient on a stretcher, is not assigned to another emergency transport and personnel resources to operate the ambulance are available. Also, we have to make sure that the closest available transporter is indeed close enough to be efficient, so we require that the approach will take no longer than the current waiting time of the patient, unless the distance to be covered is shorter than 5 kilometers. For the NAW, we allow an approach that is twice as long because it can provide more appropriate help in case of an emergency. Thus, the longer a patient is already waiting for ambulance service, the more ambulances are admissible for the transport, but their approach will take longer to reach the site of the emergency. If the patient entering the system is not an emergency, he/she is assigned to a waiting queue until a suitable ambulance is determined for the transport. The heuristic for a routing strategy that provides efficient coordination of tours while not creating unacceptable conditions for waiting patients is described below.

A transporter waiting at its parking position checks for an emergency transport every minute. If no such transport is requested, the queue of patients waiting for transportation is searched for a suitable task every $T = 15$ minutes. Note that this parameter cannot be chosen too small, as this reduces the flexibility to choose an optimal transporter for each transport. If the currently most efficient route is assigned to an ambulance after only three minutes, for example, this results in an increased mileage for the rescue organization to carry through all transports. On the other hand, for the choice of $T = 15$, a sufficiently large number of transporters and patients can be chosen from so as to optimize routes and minimize the required mileage (Koch 2002). Finally, a transport is only assigned if the approach to the closest patient's entry station is shorter than the maximum of 12 minutes and $r = 0.75$ times the current waiting time of the patient. Thus, an ambulance is assigned if the tour implies only a short approach from the parking position or if the patient has been waiting for an intolerably long period of time. The choice of the parameter r is in fact critical for the performance of the system. It turns out that a smaller value $r = 0.5$ reduces the mileage for the rescue organization, but conversely the waiting time of patients (and consequently the total time required from the request for ambulance service to the arrival at the destination) increases. This conflict of interests is discussed in Koch (2002).

When an ambulance reaches a node along the network, any pickup and drop-off actions appointed for the respective station are performed. To model this process, the transporter is delayed to allow for loading time. The duration of this delay varies stochastically according to a triangular distribution with minimum 3, mode 5 and maximum 7 minutes. This random element is also convenient because it compensates to some extent for possible modeling inaccuracies due to imprecise estimates of the size of villages or towns or neglect of the variation in the amount of traffic in larger cities. The influence of the latter factors may be seen as random effects at a similar time scale as the variation in loading time which is accounted for in the model. After loading and/or unloading patients, the planned route is updated according to the following rules:

1. If an emergency transport is being carried through or was recently assigned, the transporter moves to the next station of its route directly on the shortest path through the network. Thus, an emergency transport is inserted at the first position into the planned route and undertaken immediately.
2. If the schedule of ambulance personnel has changed and the number of operators available according to the schedule is smaller than the number actually used, no new patients are assigned to the tour, the tour is completed and the transporter moves back to its parking position and releases the operating staff. This control is realized via a comparison of the variables NR and NQ which are associated with the respective resources. The desired behavior for the release of the resource results from the definition of the resource's *capacity entity rule* as *ignore*.
3. Otherwise, the waiting queue of patients not yet assigned a transporter is searched for a possibility to coordinate any of the requested patient routes with the planned route of the transporter such that no intolerable detour results. To this aim, for every patient in the queue the data of entry station and destination are inserted into the transporter's planned tour at every possible combination of positions until an admissible tour is found. The criteria for an admissible tour in this context are:
   - The transporter's available capacity is sufficient to carry through the transport from entry station to destination, even if additional patient pickups and drop-offs are scheduled during the tour.
   - The detour for the transporter in kilometers is shorter than the maximum of 12 minutes and $r = 0.75$ times the current waiting time of the patient in minutes, but in any case less than 20 kilometers.
   - The detour for every patient assigned to the same tour of the transporter as compared with a direct transport from his/her entry position to the destination is smaller than 10 kilometers.
   - The currently planned tour contains no more than nine patient pickups and drop-offs.

   If no admissible route is found for a patient, the procedure is repeated for the remaining patients in the waiting queue.
4. If no drop-off or pickup is currently scheduled, but the capacity of ambulance personnel is sufficient to carry through further transports, the ambulance returns to its parking position, taking the shortest path but pausing at every node along the way to check for new tasks.

To illustrate our simulation model, we display a screenshot from a small demo version of our program in Figure 2. This model only contains 10 nodes from the actual network and a reduced number of ambulances and personnel for easier graphical representation. The 10 nodes of the network comprise 5 red cross stations and 2 hospitals. Currently, 1 out of 3 KTW is operating, while the 2 BKTW and 1 NAW are waiting for assignments. Consequently, 2 out of 6 currently available ambulance personnel are busy. The queue of patients not assigned an ambulance contains 2 patients, while 1 patient is waiting for pickup. The number of active transporters and of busy personnel (as compared to the personnel available due to the current duty roster) are displayed and the queues of patients not yet assigned transport or waiting for pickup are shown.

3 SIMULATION RESULTS

Our main focus was the comparison of central coordination with decentralized tour planning. In the first setting, we assume all the transporters associated with the area shown in Figure 1 to be available for ambulance service, where no restrictions on possible tours are imposed as long as patients' waiting time and total mileage for the Red Cross organization are reduced as far as possible (we already mentioned that these two aspects cannot be optimized simultaneously, however). In the latter scenario, for transports associated with the three subareas **Area 1**, **Area 2** and **Area 3** from Figure 1, only the transporters from the respective area are available to carry through the transport.

Table 1 gives the characteristic values of the system's behavior, determined for the whole region considered and additionally for the three subareas from Figure 1 with decentralized scheduling, respectively. *transfer* denotes the transfer time for each patient, *load* the time spent for pickup and drop-off, *wait* the waiting time and *TIS* the total time spent in the system. We distinguish between the values for emergencies */em* and other patients */pat*. In the simulation run, a total of 942 emergency and 13,232 other transports were carried through, with some variability in the measured characteristic values. Thus, Table 1 lists the quantities' mean values and the 95% confidence intervals.

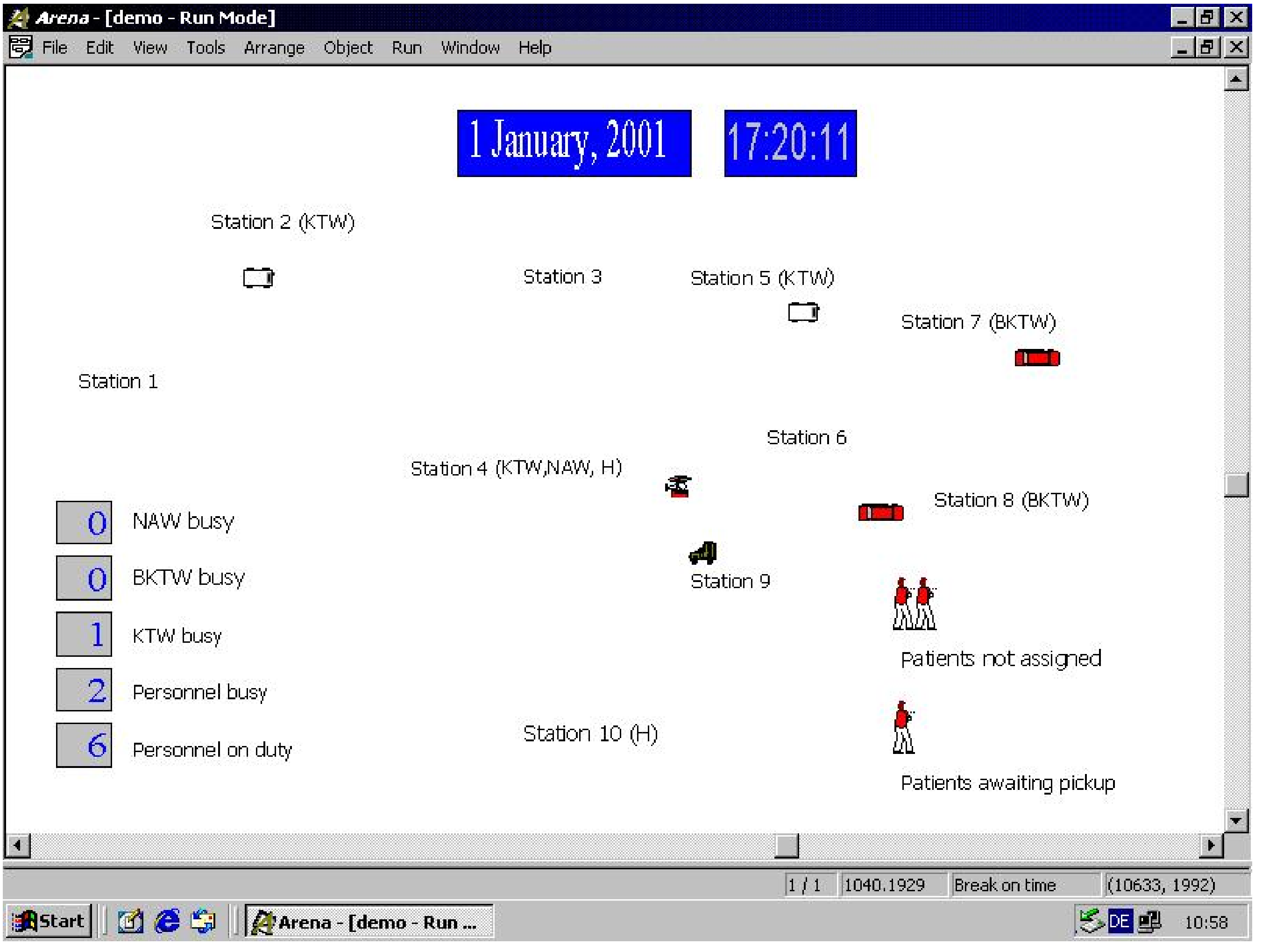

Figure 2: Screenshot from a Demo Version

The statistically significant differences between the two scenarios were determined by lumping the data from the subareas together into one data set and comparing this data set with the results for the whole region using the paired t-test at confidence level 95%.

For both the transfer and load times of emergency as well as of other patients, no statistically significant overall difference between central and decentralized coordination can be observed. However, there is apparently some slight disadvantage if **Area 1** or **Area 3** are considered separately. This is compensated for by the favorable results for **Area 2**, however.

The waiting time for patients, on the other hand, is improved by 5.29 ± 0.94 min if we consider central coordination for the whole area. This small difference can entirely be attributed to the waiting time until an ambulance is assigned. Indeed, this time factor differs by 5.09 ± 0.40 between the two scenarios, whereas the waiting time from the time a tour is assigned until the patient is picked up shows no significant difference. Note that the waiting time for decentralized planning is longest in the smaller regions **Area 1** and **Area 3**. This is obviously due to the restricted flexibility in planning if coordination is reduced to a smaller scale. There is no significant difference in the waiting time for emergencies for both scenarios.

Finally, we observe that the difference in the waiting time also has an influence on the total time a patient takes from the first request for an ambulance until drop-off at his/her destination. There is an advantage of 4.75 ± 1.19 min if central coordination is considered. Curiously, there is a statistically significant difference for emergency patients as well. The advantage of 1.88 ± 1.83 min can be neglected.

We conclude that there is a slight advantage for patients if we use central coordination instead of decentralized planning, but we are more interested in the gain in efficiency this implies for the Red Cross ambulance service. We find that central coordination reduces the total mileage by 8,346 km. This reduction by 1.19% seems rather insignificant, however. An analysis of the parameter choice (Koch 2002) indicates that apparently there is more potential gain in the

Table 1: Characteristic Values of Patient Transport

	Region	**Area 1**
transfer/em	16.68 ± 0.76 min	22.10 ± 2.13 min
transfer/pat	38.37 ± 0.59 min	46.05 ± 0.72 min
load/em	9.96 ± 0.07 min	9.81 ± 0.21 min
load/pat	13.41 ± 0.16 min	13.68 ± 0.34 min
wait/em	8.32 ± 0.69 min	10.28 ± 1.42 min
wait/pat	35.75 ± 0.76 min	55.08 ± 2.52 min
TIS/em	39.34 ± 1.27 min	49.46 ± 3.56 min
TIS/pat	87.53 ± 0.87 min	114.82 ± 2.65 min
total mileage	694,166 km	167,581 km
	Area 2	**Area 3**
transfer/em	15.64 ± 0.87 min	19.50 ± 2.64 min
transfer/pat	34.74 ± 0.67 min	39.57 ± 1.04 min
load/em	10.02 ± 0.08 min	9.98 ± 0.29 min
load/pat	13.02 ± 0.21 min	13.46 ± 0.23 min
wait/em	8.97 ± 0.92 min	11.83 ± 3.91 min
wait/pat	33.73 ± 1.45 min	48.09 ± 1.60 min
TIS/em	39.46 ± 1.66 min	48.18 ± 5.98 min
TIS/pat	81.49 ± 1.87 min	101.12 ± 2.10 min
total mileage	373,314 km	161,617 km

optimization of tour planning in every respective subarea than there is in centralization of the organization.

To prove that the small gain from centralized coordination reflects a systematic and predictable behavior of the model, we finally analyze the tours carried through in the centralized model of the whole area. It turns out that the stationing of transporters is quite adequate, and moreover we observe that special topographical features of the area under consideration influence the tours in a quite natural and predictable way. Synergies between the subareas of the region arise mainly due to special locations of the destinations of tours, and consequently we can only expect a limited gain by central coordination of tour scheduling.

The region we consider in the graph of the traffic network in Figure 1 is subdivided into three subareas associated with a number of Red Cross stations. These stations are depicted as the 11 larger dots in Figure 1. We are interested in the question whether the cooperation between these three organizational units is very strong for an optimal coordination of transports. However, this turns out to be the case only in certain subregions which can be characterized topographically in quite an intuitive way.

Let us discuss the results given in Table 2. For the three subareas, we give the number of starting points and destinations of individual patients which were transported by an ambulance from the respective area. We exclude the NAW from this discussion because it is associated with the whole region from Figure 1. “Pickup A1” – “Pickup A3” and “Drop-off A1” – “Drop-off A3” denote the total numbers of pickups and drop-offs, respectively, that were carried through in **Area 1**, **Area 2**, and **Area 3** (“A1”, “A2”, and “A3”) by ambulances associated with the respective subareas

Table 2: Analysis of the Tours

	T1	**T2**	**T3**
Pickup A1	1.350	283	9
Pickup A2	225	4.331	399
Pickup A3	5	414	1.229
Pickup H	323	4.770	532
Pickup O	3	130	13
Total	1.906	9.928	2.182
Drop-off A1	329	837	42
Drop-off A2	230	3.251	333
Drop-off A3	1	924	288
Drop-off H	1.299	4.534	1.482
Drop-off O	47	382	37
Total	1.906	9.928	2.182

(“T1” …transporters from **Area 1**, etc.). “Pickup H” and “Drop-off H” denote the same quantities for transports to and from hospitals, and “Pickup O” and “Drop-off O” refer to transports leaving the area under consideration.

Obviously, the vast majority of ambulances operate in the area they are associated with. As concerns pickup, this trend is quite distinct. There is some amount of interchange between **Area 1** and **Area 2**, and between **Area 2** and **Area 3**, but not between **Area 1** and **Area 3**. This is no surprise, as **Area 2** separates the other subareas. Moreover, **Area 1** and **Area 3** are most easily accessible via a freeway passing through **Area 2**. So especially for transports to and from hospital (the majority of hospitals is situated at the bottom corner node of the graph given in Figure 1), service of **Area 2** by ambulances from **Area 1**, and more noticeably, from **Area 3** is quite natural. Indeed, the bigger part of these transports is associated with a few places along the freeway. Very interestingly, some three or four places belonging to **Area 2** are in a special topographical situation. These places at the top right corner of the graph are quite easily accessed from the Red Cross stations in **Area 3** and are rather remote from the main part of **Area 2**. Indeed, the simulation shows that these places are served regularly by ambulances from **Area 3**.

Not surprisingly, pickup from hospital is an important factor as well. This effect is much stronger even for drop-off of patients. Still, the results show that for the remaining drop-offs, a tendency to stay in the same area can be observed. For the apparent synergies when entering a different subarea, the same factors seem to be important as in the case of pickup. The neighborhood of the freeway and the few places in **Area 2** more readily accessible from **Area 3** are those most often served by ambulances from a different subarea. Also, a big town in **Area 2** is often the destination of a transport, apparently because of many medical specialists who are resident in that town. The same town also accounted for a large proportion of the interchange to **Area 2** in the case of pickup.

It is not possible to conduct a more detailed analysis of the tours here, because we have to keep the discussion in very general terms. However, when considering the precise location of the nodes in the traffic network, it can be inferred that the tours in our simulation reflect the topographical situation very well and that ambulances keep in the area they are associated with unless a special topographical situation suggests to serve an adjacent subarea. We conclude that the layout of Red Cross stations and their association with organizational units is quite natural and efficient and moreover our simulation model works quite predictably and yields most plausible results.

To further validate the model, we compared the total mileage from our simulation with real world data. Unfortunately, only the total distances covered by all the patients, without taking into account the possibility to transport more than one patient at a time, are subject to bookkeeping at the Red Cross organization. From partial data available for some subareas, we reckon that the true value is overestimated by about 40-50% by the value on record. This value for the area under consideration gives a total of 891,139 km. Thus, the value 694,166 km from our simulation reflects the correct order of magnitude and we accept the model to work dependably.

4 CONCLUSIONS

We used an ARENA/SIMAN model to analyze possibilities to improve the logistics of ambulance service for the Austrian Red Cross rescue organization. The implementation intended a heuristic, near-optimal scheduling of patient transports. Our main aim was to compare decentralized planning with central coordination. It turned out that there is some potential to reduce waiting time for patients and mileage required to carry through the transports if the organization is centralized. However, the effect is not quite significant. Finally, we concluded that the routes taken by ambulances are in good agreement with the actual layout of organizational units of the Red Cross in the region under consideration. The association of resources with Red Cross stations seems quite natural and effective.

REFERENCES

Domschke, W. 1989. *Logistik: Transport*. 3rd ed, Volume 1. München: Oldenbourg.

Domschke, W. 1990. *Logistik: Rundreisen und Touren*. 3rd ed, Volume 2. München: Oldenbourg.

Koch, O. 2002. Discrete event simulation of the transport logistics of the Austrian Red Cross. To appear in SNE. Available at <`http://fsmat.at/~othmar/research.html`>.

Pegden, C., R. Shannon, and R. Sadowski. 1995. *Introduction to simulation using SIMAN*. 2nd ed. McGraw-Hill.

AUTHOR BIOGRAPHIES

OTHMAR KOCH is currently holding a post-doc position at Vienna University of Technology, Austria, supported by the Austrian Science Fund (FWF). He finished his Ph.D. in Numerical Analysis “sub auspiciis praesidentis rei publicae” in 1999 and has specialized in the numerical solution of singular ordinary differential equations. He was inaugurated as Univ. Doz. (Dr. habil.) in Numerical Analysis in 2003. In 2002 he carried through the project described in this paper as part of his civil service with the Austrian Red Cross.

HEINZ WEIGL is president and founder of ESLA in Austria, operating in the field of logistics systems and simulation. He is also lecturer at the University of Technology Vienna (TU-Wien), University of Technology Dresden, University of Economics Vienna and Fachhochschulen in Austria. He received his Dipl.-Ing. in Electrical Engineering in 1980 from TU-Wien, his Dr. techn. in 1992 from TU-Wien and his Dr. Ing. habil. (production logistics) from the University of Technology Dresden in 2001.