SEARCHING FOR THE BEST LOCATIONS FOR SERVICE FACILITIES ALONG A FREEWAY

We-Min Chow and Adolf D. May, Operations Research Center,
University of California, Berkeley

Two models were used to find the optimal locations for service facilities along a freeway. The first one is a simulation model called FREEQ. For a given accident or incident on the freeway, FREEQ can be employed to generate all necessary information, such as total travel time and individual average travel time on the freeway, provided that the demand pattern and the physical configuration of the freeway are known. Based on these results, an optimization model is used to search for the best locations for service facilities so that the total delay time caused by the accident or incident or the response time of the service unit is minimized. The Eastshore Freeway in the San Francisco Bay area was chosen to be the study area. Thus, a numerical problem is also given.

MAJOR PROBLEMS may develop due to the occurrence of accidents or incidents on freeways: Traffic flow will be interrupted because of reduced capacity, thereby causing traffic congestion and delays to passing motorists. Also, the waiting time for necessary service can possibly be vital to the survival of stranded motorists. Consequently, the objective should be to minimize total delay and response time (the time until service vehicles reach the accident location).

The purpose of this paper is to search for the best locations of service facilities along a freeway. Basically, two models are presented: a simulation model called FREEQ, which will be used to generate all necessary data such as average individual travel time and total travel time, and an optimization model employed to find the best locations of the service facilities based on the results generated from FREEQ. A schematic model is shown in Figure 1.

ASSUMPTIONS AND PROCEDURES OF SIMULATION MODEL

This simulation model was first developed by Y. Makigami, L. Woodie, and A. D. May in August 1970. A FORTRAN IV computer program written for a CDC 6400 is available.

The basic assumptions of the model are as follows:

1. Traffic is treated as a compressible fluid where an individual vehicle is regarded as an integral part of the flow and is considered individually;
2. Within a given time interval (usually 15 min), traffic demands remain constant and do not fluctuate over that time interval, and, for given subsections, traffic demands are expressed as a step function over the entire time period under consideration;
3. Once traffic demands are loaded onto the freeway, the demands propagate downstream instantaneously unless there are capacity constraints; and
4. Capacities of subsections, including weaving sections and merging points, are estimated by using Highway Capacity Manual methods (3).

If both the physical configuration of the freeway and the traffic demand pattern are known, then freeway performance can be evaluated by this model.

Publication of this paper sponsored by Committee on Motorist Services and Committee on Freeway Operations.

The stepwise procedure of the model is as follows:

1. Read input data, which consist of number of subsections, number of lanes in each subsection, capacity of each subsection, length of each subsection, truck factor in each subsection, type of ramps in each subsection (e.g., on, off, left, multilane), ramp capacity, and origin-destination demand pattern for each 15-min time slice.
2. Compute the demand.
3. Modify O-D distribution (ramp analysis).
4. Modify the freeway capacity (weaving analysis).
5. Compare the demand and the capacity (if demand > capacity, go to 6; if demand ≤ capacity, go to 7).
6. Go through queue increasing process; calculate the average speed, travel time, queue length, and travel distance. Then go to 9.
7. Check whether there is any queue remaining from the last 15-min time slice. If there is, go to 8. Otherwise, go through nonqueuing process, and go to 9.
8. Go through queue discharging process; calculate the same things as in step 6.
9. Print out the results.

This completes a whole cycle for each time slice. Some of the current results such as queue length and number of vehicles in the queue are used as the initial condition for the next computation cycle (the next time slice).

The basic idea used in the FREEQ model is to divide the freeway into subsections according to its physical configuration, so that each subsection can be treated as a uniform pipe and the capacity of every point in a specified subsection is always the same. Whenever an accident (inasmuch as there is no need to distinguish between accident and incident, the term accident will be used throughout this paper) occurs, the capacity is reduced until the disabled vehicle is removed. If the blockage time (during this time interval, the freeway may operate at reduced capacity) and the effective length (length of the freeway segment having reduced capacity) are both known, the modification can easily be made by subdividing the time slices and subsections into smaller intervals. Therefore, the uniformity property in each new subsection is preserved, and FREEQ can be used directly. Of course, in this situation some of the time slices may be less than 15 min in duration.

RESULTS AND ANALYSIS

The present application of this study is limited to the northbound Eastshore Freeway in the San Francisco Bay area. This freeway is composed of 30 subsections (in case of no accident). The time interval covers a 2½-hour afternoon peak period from 3:45 to 6:15 p.m.; hence, there are a total of 10 time slices.

Normal Case: No Accident

Figure 2 shows the speed, density, and queue length in a time-distance space. The traffic volume in each subsection over each time interval can be computed by the relation $q = \mu k$; i.e., volume (vph) = speed (mph) × density (vpm).

Subsection 20 becomes a bottleneck at the beginning of the second time slice, and subsections 5 and 25 become bottlenecks at the beginning of time slices 4 and 6 respectively. The shock wave is recovered in time slice 7. The total travel time TTT = 5,017 passenger-hours.

Accident Case

In this report only the single-accident case has been considered, and 16 of the 30 subsections have been investigated for this example. These 16 subsections were chosen from the Gilman on-ramp to the San Pablo Dam Road off-ramp, inasmuch as this region covers all possible traffic situations such as forming and recovering of shock waves, bottlenecks, and congested flow and free-flow cases. Moreover, this region is far from the beginning of the main-line freeway, and, if an accident occurs, the chance that the queue backs up out of the upstream boundary of the freeway is small. The accident may occur in only 16 of the 30 subsections, but the total travel time over all 30 subsections is studied. It is assumed that

1. Only one accident can happen in the peak period,
2. When the accident occurs, one lane of capacity is lost,
3. The effective length of the accident is 100 ft, and
4. The accident spot is located at the midpoint of the subsection (and also at the midpoint of 100-ft section).

The last assumption will approximately give the average value of each measurement (total travel time, average individual travel time in each subsection, and so forth). In other words, when an accident occurs in a specified time slice and subsection, the average measurement is approximately equal to the computed measurement as if the accident took place at the midpoint of the subsection.

Because there are 16 subsections and 10 time slices, 160 accidents were generated. Each accident corresponds to an (s, t) pair if it occurs in subsection s during time slice t. The delay time of each accident is defined by

$$\Delta TTT(s, t) = TTT(s, t) - TTT$$

That is, delay = total travel time when an accident occurs in (s, t) - total travel time when there are no accidents.

In Figure 3, the number in each cell is the average delay in passenger-hours when an accident occurs in the corresponding subsection and time slice and blocks the traffic for half an hour. The blank cells show that the delay time due to the accident is less than 10 passenger-hours. Comparing this with 5,017 passenger-hours (TTT under normal condition) reveals that the increment is less than 0.2 percent and hence is considered to have no effect.

Clearly, it can be seen from this figure that, if an accident happens in time slice 9 or 10, it does not interrupt the traffic flow very much. According to Figure 2 in these two time slices, the freeway has low density and high speed. This implies that the traffic load is light. Consequently, if one lane of capacity is lost, the traveling speed will not be affected too much.

It is also interesting to note that an accident may have little effect if it occurs in the normal congestion area. In this case, the bottleneck is shifted upstream, and the traffic condition in the area downstream of the accident location is improved. Data shown in Figure 3 demonstrate that, if an accident takes place at the normal bottleneck, the delay time is significantly high, but, if it occurs in the congestion area (i.e., upstream of the bottleneck), the accident will cause little delay.

OPTIMIZATION MODEL

The best location for service facilities will minimize the maximal possible total delay time or minimize the maximal possible travel time of a service vehicle. In any case, a good emergency system should be able to clear off the accident promptly, that is, to minimize the blockage time. In general, the blockage time can be divided into three nonoverlapping parts: detection time, waiting time for service or, equivalently, the response time of a service vehicle, and on-site service time.

The detection time is dependent on the detection system, whether electronic detectors, call boxes, emergency telephones, patrolling vehicles, helicopters. The on-site service time, on the other hand, is contingent on the type of accident. The controllable variable in our problem is response time, and the investigation will be undertaken by assuming different values of the sum of detection time and on-site service time. The problems are solved by first selecting certain upper limits in total delay time or response time and then finding "admissible" locations for service facilities so that these limits will not be violated. By changing these upper limits, new problems can be formulated. Therefore, our work is to solve a sequence of parametric optimization problems, each of them corresponding to a specified upper limit value.

Figure 1. Schematic of FREEQ model.

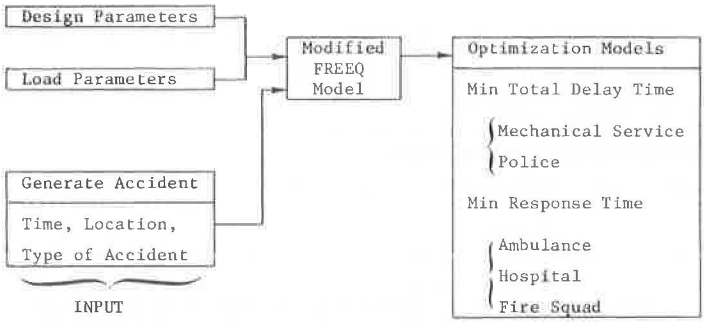


Figure 2. Speed, density, and shock wave diagram.

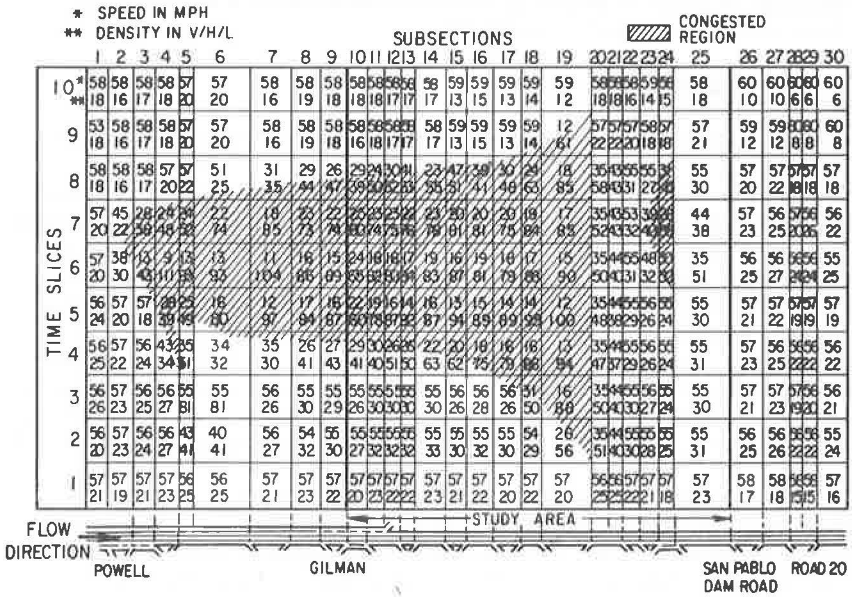
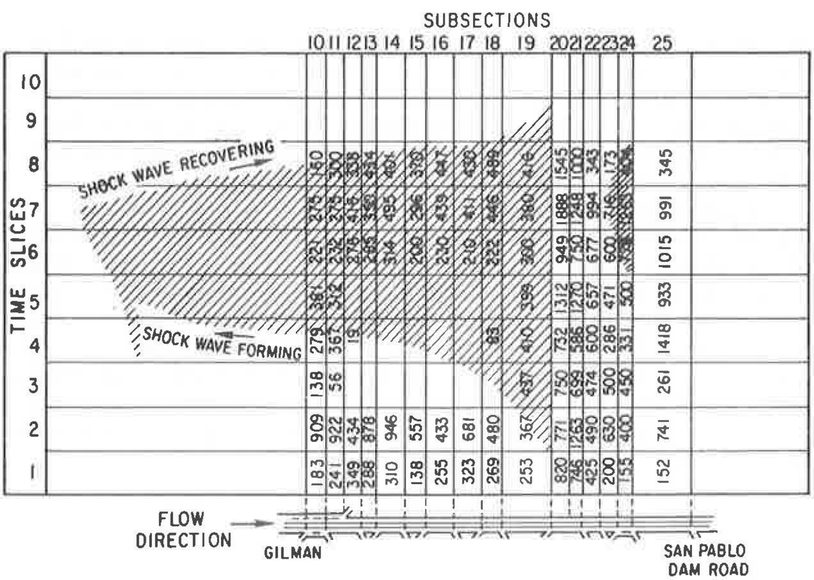


Figure 3. Delay in passenger-hours due to accident occurrence.



Mathematical Formulation

For each (s, t) pair, it is assumed that the total delay time to freeway users is a linear function of blockage time of the accident. Because the min-max criterion is employed, for each subsection s only the most critical time slice needs to be considered. (During this time slice, if an accident occurs, the delay time takes on its maximal value. For example, in Figure 3, for the last subsection, the most critical time slice is 4 where an accident causes a delay of 1,418 passenger-hours.) If the locations of service facilities have been found in this way, then the same solution will automatically satisfy the constraints for the other time slices.

Based on this assumption, the total delay due to the (s, t) accident is given by

$$\Delta TTT(s,t) = TTT(s,t) - 5{,}017 = B(s,t) \times DT$$

where

DT = blockage time of (s, t) accident and
B(s, t) = rate of contribution of (s, t) accident.

For each s, the most critical time slice t^* will be one in which

$$\bar{B}(s) = B(s,t^*) = \max_t B(s,t)$$

or, equivalently,

$$TTT(s,t^*) = \max_t [TTT(s,t)]$$

Now, the problem is to investigate a number of potential locations and to find the best ones. Let

$$x_i = \begin{cases} 1, & \text{if a service facility is needed at location } i \\ 0, & \text{otherwise} \end{cases}$$

Then the problem is given by

$$\text{Minimize } Z = \sum_{i=1}^{n} c_i x_i$$

subject to $TTT(s,t^*) \le SL$, for all s, and $x_i = 0$ or 1, for $i = 1, \ldots, n$, where

c_i = the cost incurred if a service facility is established at the i th candidate location and
SL = preselected upper limit (or service level).

Note that, first, $TTT(s,t^*) \le SL$, for all s, implies that $TTT(s,t) \le SL$ for all s and t and, second, $TTT(s,t^*)$ is a function of x_i's. This value can be evaluated based on the result from the simulation model.

To solve the problem, two things must be predetermined: the desirable service level SL and the sum of the detection time and on-site service time TT. If a service vehicle is dispatched from the i th service facility to an (s,t^*) accident location, the travel time of this vehicle $Tx_i(s)$ will be evaluated, based on the result from our simulation model. The duration of the accident DT is simply the sum of TT and $Tx_i(s)$; therefore,

$$TTT(s, t^*) = DT \times \bar{B}(s) + 5{,}017$$

$$= [TT + Tx_i(s)]\bar{B}(s) + 5{,}017$$

and the problem can be reformularized as

$$\text{Minimize } Z = \sum_{i=1}^{n} c_i x_i$$

subject to $Tx_i(s) \leq \frac{SL - 5{,}017}{\bar{B}(s)} - TT$ for all s and x_i, and $x_i = 0$ or 1, for $i = 1, \ldots, n$.

This type of problem can be solved in several ways, such as integer programming, dynamic programming, and branch and bound procedure. However, only the following method will be used in this paper.

Algorithm

A minimal set solution (MSS) is a set that is a solution but no proper subset of it can be a solution. For instance, if there exist set solutions {1, 2}, {1, 3, 4}, {1, 2, 3}, and {1, 2, 3, 4}, then only {1, 2} and {1, 3, 4} are MSSs because $\{1, 2\} \subset \{1, 2, 3\} \subset \{1, 2, 3, 4\}$.

The algorithm used here is to search for a sequence of MSSs. If {i, j, k} is the MSS, it indicates that, when the service facilities are established at the candidate locations i, j, and k, the desirable service level can be attained.

Initially, the problem is formularized as a matrix. In the following example, a 1 is put into the cell (i, j) if location i can provide the service for a (j, t^*) accident without violating the preselected SL limit. Then the problem is solved by searching for all MSSs from the incident matrix, and only one column is considered at a time. The stepwise procedure of the algorithm is given as follows:

1. Find all MSSs for column one and treat these solutions as the current partial solution. In the example, they are {1} and {2}.
2. Find all MSSs for the next column.
3. Combine the partial solution and the solution just obtained to find all MSSs again. Treat the result as the current partial solutions.
4. Repeat steps 2 and 3 until all columns have been considered. At this moment the completed solutions are determined, which are the desirable MSSs for the whole matrix.

Example

Candidate Locations i	Subsections j: 1	2	3	4	5	6
1	1	1	1			
2	1	1	1	1	1	
3		1	1	1		
4			1		1	1

Stage 1—Initial partial solution (from column 1): {1} and {2}.

Stage 2—MSSs for column 2: {1}, {2}, and {3}. Combine stages 1 and 2; the admissible partial solution will be $\{1\} \cup \{1\}$, $\{1\} \cup \{2\}$, $\{1\} \cup \{3\}$, $\{2\} \cup \{1\}$, $\{2\} \cup \{2\}$, and $\{2\} \cup \{3\}$. Because $\{1\} \subset \{1, 2\}$, $\{1\} \subset \{1, 3\}$, and $\{2\} \subset \{2, 3\}$, the current partial solutions should be {1} and {2}, which are also the MSSs for columns 1 and 2.

Stage 3—MSSs for column 3: {1}, {2}, {3}, and {4}. Admissible partial solutions: {1}, {2}, {1, 3}, {1, 4}, {2, 3}, and {2, 4}. Current partial solutions: {1} and {2}.

Stage 4—MSSs for column 4: {2} and {3}. Admissible partial solutions: {1, 2}, {1, 3}, {2}, and {2, 3}. Current partial solutions: {1, 3} and {2}.

Stage 5—MSSs for column 5: {2} and {4}. Admissible partial solutions: {1, 2, 3}, {1, 3, 4}, {2}, and {2, 4}. Current partial solutions: {2} and {1, 3, 4}.

Stage 6—MSS for column 6: {4}. Admissible partial solutions: {2, 4} and {1, 3, 4}. Current partial solutions: {2, 4} and {1, 3, 4}.

Inasmuch as all the columns have been scanned, the MSSs for the matrix will be {2, 4} and {1, 3, 4}. This means that the service facilities should be established either at locations 1, 3, and 4 or at locations 2 and 4. If the incurring cost for each location is known, the decision can be made by simply comparing the total cost for each MSS.

Clearly this method can find all the alternative solutions. If the problem is to minimize the maximal possible response time of the service vehicle, the same method is still applicable. A 1 is put into cell (i, j) if station i can send a service vehicle to the accident location within the preselected time limit. In this case it is not necessary to know the detection time and on-site service time.

It is also clear that, because no calculation is needed, the chances of making errors are reduced. When the preselected limit is changed, the post-optimality problem is easy to handle. To see this, first the problem is solved by using a large preselected limit (the desirable SL on the freeway). As this limit is decreased, some of the columns will have a different structure; i.e., the number of 1's is reduced. Those columns can be treated as if they were the m + 1, m + 2, and so on (if initially there are m subsections), and a post-optimality problem will be solved with current partial solutions equal to the previous MSS. This procedure is legitimate because any MSS generated from the new columns can satisfy the corresponding columns in the original incident matrix. In the preceding example, if the 1 in cell (3, 4) is removed, then the new column generates MSS {2}. Combining this to previous solutions {2, 4} and {1, 3, 4}, the final MSS will be {2, 4} only (because $\{2, 4\} \cup \{2\} \subset \{1, 3, 4\} \cup \{2\}$).

Algorithm Refinement

When the problem size is increased, the effort of determining MSS may grow very rapidly. It is desirable to go through a certain elimination procedure to cut down the size of the problem. Some refinement is therefore given.

1. If for each 1 in a given column there is also a 1 at the corresponding position in another column, then the latter can be removed from the matrix and never needs to be considered because any solution of the first column is a solution of the second column.
2. If for each 1 in a given row another 1 can be found at the corresponding position in some row, then the latter dominates the former. Physically, it means that there is another service facility that can provide the same service as the given facility. Hence, the dominated one can be neglected unless it has a smaller incurring cost than the other does.
3. If there is a column that has a 1 in every row, any facility can render the necessary service to this subsection. Hence, this column can be eliminated from further consideration.
4. If a column contains a single 1 in the ith row, then the candidate location i must be in every MSS. Consequently, any other columns that have a 1 in the same position can be removed from further consideration.

In the previous example, it can be seen that

1. Column 1 is a subset of column 2, thus eliminating column 2;
2. Column 3 can be dropped, because any choice of the service facilities will be the solution with respect to this column; and
3. Column 6 contains a single 1; therefore candidate location 4 must be in the MSS, and columns 3 and 5 do not have to stay in the matrix.

The equivalent matrix is now

$$
\begin{array}{ll|c|c|c|}
 & & \multicolumn{3}{c}{\text{Subsections } j} \\
 & & 1 & 4 & 6 \\
\hline
 & 1 & 1 & & \\
\text{Candidate Locations} & 2 & 1 & 1 & \\
i & 3 & & 1 & \\
 & 4 & & & 1 \\
\hline
\end{array}
$$

The final MSSs are clearly {2, 4} and {1, 3, 4}, which is of course consistent with the previous result.

If the incurring cost $c_1 \geq c_2$, row 1 is a subset of row 2 and therefore can be eliminated from further consideration. The new equivalent matrix will be the same as the above one except that the first row is removed. Consequently, the only MSS is {2, 4}. Because $c_1 + c_3 + c_4 > c_2 + c_4$, the solution {2, 4} has the minimal cost.

For most of the practical problems, one would expect that the problem size can be significantly reduced by going through the elimination procedure.

SEARCHING FOR THE BEST LOCATIONS

Procedure Used to Minimize the Maximal Possible Delay

The following general procedures are given:

1. Obtain the input data for **FREEQ**.
2. Generate the accident systematically for each (s, t) pair.
3. Run modified **FREEQ** to get TTT(s, t) and the average travel time in each subsection for every (s, t) accident.
4. Find the most critical time slice for each subsection.

$$\text{TTT}(s, t^*) = \max_t [\text{TTT}(s, t)]$$

5. Calculate the rate of change of delay $\bar{B}(s)$ for each (s, t^*).

$$\bar{B}(s) = [\text{TTT}(s, t^*) - \text{TTT}]/\text{DT}$$

6. Choose the potential locations for service facilities.
7. Compute the response time $T_x(s)$ of the service vehicle from each of these potential locations to the most critical accident locations.
8. Select the value of the sum of detection time and on-site service time TT, and compute the contribution of the blockage time to the total delay.

$$\text{SL}_x(s) = [\text{TT} + T_x(s)] \times \bar{B}(s)$$

9. Select the desirable SL, and construct the incident matrix. If $\text{SL}_x(s) \leq \text{SL}$, then put a 1 in the cell (i, j) where i = x and j = s.
10. Search for the MSSs from the incident matrix.
11. Change TT and SL. Find the corresponding minimal set solutions by solving the post-optimality problem.

Procedure Used to Minimize the Maximal Possible Response Time

The first three steps are exactly the same as in the previous procedure.

4. Calculate the response time $T_x(s, t)$ from location x to the (s, t) accident location for all x, s, and t.
5. Find $\bar{T}_x(s) = \max_t T_x(s, t)$ for all x and s.

6. Select the upper limit for the response time, called T.
7. Prepare the incident matrix. If $\bar{T}_x(s) \le T$, then put a 1 in the cell (i, j) where i = x and j = s.
8. Search for the minimal set solutions from the incident matrix.
9. Change the value of T, and solve the post-optimality problem.

Numerical Solutions of Minimal Delay Time

The value of TTT(s, t) is evaluated by using FREEQ with the duration time DT equal to 30 min. The results are shown in Figure 3. Meanwhile the average individual travel time in each subsection is obtained for every (s, t) accident. The values of $\bar{B}(s)$ are as follows:

s	$\bar{B}(s)$	s	$\bar{B}(s)$
1	30.3	9	16.3
2	30.7	10	14.6
3	14.5	11	62.9
4	29.3	12	42.3
5	31.5	13	33.1
6	18.6	14	23.9
7	14.9	15	42.1
8	22.7	16	47.3

Six potential locations will be selected, each located near the on-ramp. The traveling speed in the opposite direction of traffic flow (i.e., from downstream to upstream) is assumed to be 35 mph; therefore, the response time of the service vehicle is ready to be evaluated. The travel time on the freeway is of course obtained from the result of FREEQ. (Recall that this travel time corresponds to the accident case in the most critical time slice.) The estimated values of response time of a service vehicle from location x to subsection s is given in Table 1. Table 2 gives all values of the contribution of response time to the freeway delay. For a given value of TT, the contribution of the blockage time to the delay on freeway can be computed as $\bar{B}(s)\ [T_x(s) + TT]$ for all x and s. Consequently, if the desirable service level is chosen, the incident matrix can be obtained. By using the method suggested previously, it is easy to find the best locations for the service facilities.

The optimal solution curve is shown in Figure 4, the horizontal scale is the value of the sum of detection time and on-site service time TT, whereas the vertical scale indicates the total delay time in passenger-hours. The lower right side of the curve is the infeasible region, and the upper left side of the curves is the feasible region. In the figure, four solution curves are shown that correspond to the solutions {1, 2, 5, 6}, {2, 5, 6}, {5, 6} and {5} respectively. Hence, for a given value of TT, the maximal possible delay of an accident can be found from the graph for each solution.

Numerical Solutions of Minimal Response Time

Because the computation procedure is almost the same as the preceding section, only the result will be given here.

Solutions	Maximal Possible Response Time (min)
{1, 2, 3, 4, 5, 6}	2.46
{1, 2, 4, 5, 6}	3.59
{2, 3, 5, 6} or {2, 4, 5, 6}	3.82
{2, 5, 6}	4.69
{5, 6}	5.21
{5}	8.29

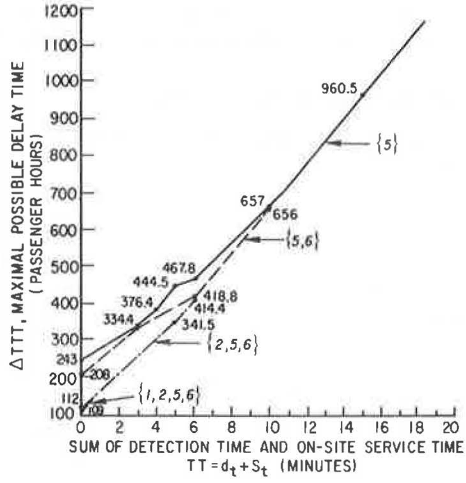
Table 1. Response time (in min) from service station to accident in subsection during the most critical time period.

	Service Station Location					
Subsection	1	2	3	4	5	6
16	23.56	16.94	13.71	8.22	3.89	2.31
15	25.58	21.84	18.01	12.90	4.95	0.24
14	16.85	14.07	10.52	7.18	2.32	3.85
13	18.32	15.25	11.96	7.79	1.74	3.27
12	28.40	22.92	17.83	12.06	1.10	2.63
11	27.58	23.45	17.44	10.71	0.43	1.96
10	5.42	4.57	3.70	2.38	4.09	5.62
9	9.64	7.03	3.82	0.65	2.36	3.89
8	6.15	4.69	2.23	3.59	5.31	6.84
7	7.13	4.27	0.76	2.12	3.84	5.37
6	3.67	2.04	3.43	4.79	6.51	8.04
5	3.64	1.00	2.39	3.75	5.47	7.00
4	2.46	3.82	5.21	6.57	8.29	9.82
3	1.71	3.07	4.46	5.82	6.54	8.07
2	1.02	2.38	3.77	5.13	6.85	8.38
1	0.38	1.74	3.13	4.49	6.21	7.74

Table 2. Contribution of response time to total delay time on freeway (in passenger-hours).

	Service Station Location					
Subsection	1	2	3	4	5	6
16	1,125	825	649	388	184	109
15	1,075	920	762	543	208	101
14	428	336	252	154	55	92
13	675	505	397	258	58	108
12	1,205	972	755	512	47	113
11	1,770	1,475	1,095	659	27	126
10	79	67	54	35	60	82
9	157	115	62	11	38	64
8	140	106	51	82	121	155
7	106	64	11	32	57	80
6	68	38	64	89	121	149
5	115	32	75	118	155	221
4	72	112	153	193	243	288
3	25	44	65	84	95	117
2	31	73	116	157	210	257
1	12	53	95	136	188	234

Figure 4. Optimal solution curve.



DISCUSSION OF MODEL

One assumption that has been made is the linear relationship between the duration time of the accident and the total delay to freeway users. This may not be always the case. If freeway operation is interpreted as a queuing system so that the demand is considered as arrivals to the system and freeway capacity as the service rate, then during the peak period this linear approximation usually will give a fairly good result. This assumption is made merely for simplifying the calculations. As a matter of fact, the problem can still be solved without using the assumption. The modified FREEQ model can be used to find the relationship between the duration time of the accident and the delay time to freeway users, and the rate of change of delay is then a function of time. For different values of TT and $T_x(s)$, the magnitude of $SL_x(s)$ can be determined as a function of $DT = [TT + T_x(s)]$. Consequently the incident matrix is capable of being constructed.

In formulating the problem, the constraints are established by first selecting TT and the desirable SL and letting

$$SL_x(s) = [TT + T_x(s)] \times \bar{B}(s) \leq SL - 5{,}017 \qquad (1)$$

where $\bar{B}(s) = \max_t B(s,t)$ for all s, and $T_x(s) = T_x(s,t^*)$ such that $B(s,t^*) = \bar{B}(s)$. But the actual constraints should be

$$[TT + T_x(s,t)]\, B(s,t) \leq SL - 5{,}017$$

for all s, t, and x where $T_x(s,t)$ is the travel time of the service vehicle from location x to an (s, t) accident location. This is equivalent to

$$\max_t \{[TT + T_x(s,t)]\, B(s,t)\} \leq SL - 5{,}017 \qquad (2)$$

for all s and x. Clearly, the two sets of constraints (Eqs. 1 and 2) may not be the same, inasmuch as an accident that requires the longest response time may not delay the total travel time most. It is clear that

$$[TT + T_x(s)]\, \bar{B}(s) \leq \max_t \{[TT + T_x(s,t)]\, B(s,t)\}$$

for all s and x. If the blockage time DT is large enough, then

$$\bar{B}(s) \times DT \geq B(s,t) \times DT$$

for all t and

$$[TT + T_x(s)]\, \bar{B}(s) = \max_t\, [TT + T_x(s,t)]\, B(s,t)$$

for all x and s, and the constraints (Eqs. 1 and 2) are identical. In our problem, this is the case when

$$DT = TT + T_x(s) \geq 5 \text{ min}$$

and this is of course a relevant assumption for the realistic cases.

Although the present problem has been solved with Eq. 1, the same technique can be used if Eq. 2 is employed.

The present study is concerned with one-direction traffic flow; a more realistic result should be obtained by considering two-way traffic either along a given freeway or within a specified network of freeways. This is simply a generalization of the present work. The same procedures and techniques are still applicable.

FUTURE RESEARCH

Three major studies are considered as future research in this field.

Investigation of Multiaccident Case

It is not impossible that during a short time period there is more than one accident found on the same freeway. The problem will be to determine (a) the relationship between delay and duration time of the accidents and (b) the optimal locations for the service facilities.

Finding the Optimal Number of Service Vehicles

Two things are involved in this problem. First, the probability distribution of accident over the time space must be determined. Second, the type of accident and the duration of the service time including response time, on-site service time, and return of the service units should be determined. Perhaps the best that can be done for this type of problem is to find the confidence level for each preselected waiting time of stranded motorists for necessary service.

Finding the Best Locations for Service Facilities for the Future Time Period

From the present result, it can be seen that the traffic demand pattern does affect the solutions of the problem. If the demand is changed, the current solutions may not be optimal anymore. It would be interesting to know how the changes in demand pattern can affect the present result and what the optimal solutions should be for a given future period. To answer these questions, the very first study should be to investigate the stochastic property of traffic and future traffic demand.

ACKNOWLEDGMENT

This research has been supported by the National Science Foundation with the Operations Research Center, University of California, Berkeley.

REFERENCES

1. Keller, H. An Analysis of Freeway Emergency Service Systems. Operations Research Center, Univ. of California, Berkeley, ORC 69-20, Aug. 1969.
2. Sakashita, M., Lu, C. K., and May, A. D. Evaluation of Freeway Emergency Service Systems Using a Simulation Model. Operations Research Center, Univ. of California, Berkeley, ORC 70-23, July 1970.
3. Highway Capacity Manual. HRB Spec. Rept. 87, 1965.
4. Makigami, Y., Woodie, L., and May, A. D. Bay Area Freeway Operation Study—Final Report, Analytic Techniques for Evaluating Freeway Improvements, Part I of III: The Freeway Model. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Aug. 1970.
5. Toregas, C., Swain, R., Revelle, C., and Bergman, L. The Location of Emergency Service Facilities. Operations Research, Vol. 19, No. 6, Oct. 1971, pp. 1363-1373.
6. Hu, T. C. Integer Programming and Network Flows. Addison-Wesley, 1969.

DISCUSSION

Robert L. Hess, University of Michigan

The paper presents a description of the modified FREEQ simulation model and an optimization model. The modified FREEQ simulation accepts all of the normal data associated with design and demand in terms of length segments and time slices with the special capability of being able to further subdivide length segments and to increase time slices according to the specifics of an assumed blockage. Flow in each original or modified subsection of length is taken as a compressible fluid in a uniform pipe. In the present study, two extremely important assumptions are made: that an accident is equivalent to loss of one lane's share of capacity and that the subsegment of length around the accident is fixed.

The output of FREEQ is apparently more influential on the final results than are the different choices of optimization procedure. The model chosen appears to have excellent clarity and benefits from being rather easily manipulated by hand operations by the user in simple cases to gain understanding, familiarity, and trust in the procedure.

The reviewer is prompted to question, however, the influence of the assumptions made in modified FREEQ on the final outcome of the optimization procedure. Basic to this question is the understanding that the freeway capacity submodel is not adequate for subsections where the number of lanes changes. If this is true, then is it admissible to simply cut a subsection with a blocked lane into three smaller subsections, the first and third of original capacity and the second being characterized by, say, $\frac{2}{3}$ capacity? The basic question is, What impedance is represented by the reduction or increase of one lane? Second, it would appear that an accident in one lane would introduce a weaving section that was not present before the accident and that its existence would have an effect on the capacity of a section of the subsegment prior to the accident location. Finally, the estimation of average speed from the relationship between the volume-capacity ratio and the operating speed shown in the Highway Capacity Manual appears questionable. Experience in fitting freeway capacity models to actual data seems to indicate that drivers, even under normal conditions, obtain actual speeds that differ from those estimated from the Manual. Furthermore, experience in accident investigation indicates that drivers, once perturbed in speed due to congestion, behave differently if the cause of the congestion is an accident than, say, if it were a maintenance operation.

It would seem that these questions do not speak to the real contribution of the current paper, i.e., a procedure for minimizing the maximal possible delay times of motorists or of response time of a service unit given an adequate flow model. Still, to the extent that modified FREEQ might not in fact adequately simulate the traffic situation, use of the minimization may suffer. To turn the questions around would be to ask, How sensitive is the result of the minimization process to variations in input? That is a question that was not addressed in the paper.

The paper concludes with suggestions for future research. Each of the projects mentioned would appear mathematically feasible but could potentially outstrip the ability of a field group to provide model validation. This discussant suggests that the authors might wish to frame a simpler problem that could have physical fruition sooner. The suggested project is to relate the optimal service station location to blocks of time suggesting that, by shifting the location during a 24-hour period, the system might achieve a lower maximum than expected. How would the maximums vary if a given station could have 2 or 3 possible different locations on a 12- or 8-hour shift basis?

Everett C. Carter, University of Maryland

The authors have presented the results of rather extensive modeling efforts at locating motorist service facilities for freeways. In general, what appears to be a very workable methodology for determining the location of motorist service stations has been developed and documented. As a result, a two-stage model evolved with two techniques presented for solving the second model. The first model, a simulation model, appears to operate well except that only a single incident (or accident) is generated for each run, whereas there is some probability of multiple incidents during a peak period for the average urban freeway. Although the assumption of a fixed demand pattern for a 15-min time slice is generally reasonable, short time (on the order of 5 min) fluctuations in demand may occur. It would be interesting to test the sensitivity of this assumption by running the simulation model for 5-min time slices.

The assumption made on the linear relationship between the duration of an incident and total delay on the freeway may not always hold. This is recognized by the authors later in the paper and in their proposed future research. It would seem that, for two incidents of equal duration, the demands might be substantially different, resulting in a significantly longer queue of delayed vehicles in one case. Hence it is expected that the total delay time required to discharge the queue on the freeway would be different for the two incidents of equal duration.

Also, proposed future research to consider the multiple-incident case should be encouraged. It would appear that the FREEQ simulation model could be modified to yield a time-space distribution of incident occurrence that is related to some actual observed probability distribution, reflecting geometrics and other characteristics. The total delay or TTT would likely be quite high in the case of multiple incidents in one time slice or in the case of a second incident occurring in the region affected by the first before the delay (queue) from the original incident had dissipated.

The total delay to freeway traffic due to an incident depends on a very complex set of factors, which is recognized by the authors. Such factors include

1. The type of detection;
2. Surveillance and/or motorist service (e.g., roadside communication system) system;
3. Detection time, which would be influenced by patrols, spacing of communication devices, and the like;
4. On-site service time, which could be described by a distribution that varies with the type of incident and the number and type of vehicles involved;
5. Time required for the service vehicle(s) to respond and reach the incident site, which varies with the traffic flow (or time of day) and travel distance; and
6. Physical factors such as geometrics and adequacy of shoulders.

The assumption by the authors of an average duration time of 30 min with an effective blockage length of 100 ft appears to be a reasonable value for representing this complex situation.

Because the paper does not contain an explanation of how the capacities of the subsections are estimated, it is not possible to judge the adequacy of the travel time estimates (obtained from the Highway Capacity Manual curve of speed versus volume-capacity ratio), which are the basis for the optimization model. Also, because the queue increasing and discharging processes, which are critical to estimating passenger hours of delay, are not explained, it can only be assumed that these steps in this simulation model adequately represent traffic flow on a freeway when an incident occurs.

A model that uses the results of the simulation model to simultaneously determine locations (and number of emergency vehicles) for motorist service stations and minimize the cost of providing such service would be desirable. Of course, constraints on level of service (in terms of total travel time) should be adhered to. In addition, constraints on response time should be employed, especially for emergency medical service needs. In fact, minimum response time will probably be attained by separate response vehicles for medical needs and mechanical needs, which probably involves two different sets of response times, one for each vehicle type. As indicated by the authors, the response time for mechanical needs may be directly reflected in delay or duration of the incident. However, it would be desirable to use an optimization model that included constraints on medical service response times. It is recognized that this will add complexity to the model.

The optimization model developed by the authors minimizes the cost of establishing and operating service stations and includes a constraint on maximum total travel time (service level). This model can be expanded to the broader concept expressed above by using cost functions that reflect the total cost per vehicle response for medical and mechanical responses. The objective function could take the following form:

$$\text{Min } Z = \sum_{i=1}^{n} \sum_{s=1}^{m} C_{is}^{1} X_{is}^{1} + \sum_{i=1}^{n} \sum_{s=1}^{m} C_{is}^{2} X_{is}^{2}$$

subject to TTT $(s, t^*) \leq SL$ and $T_x(s, t) \leq T^*$ for all X^1_{is} where

C^1_{is} = total cost per vehicle response for emergency medical needs from service station at i to section s,
C^2_{is} = total cost per vehicle response for mechanical demands from service station at i to section s,
X^1_{is} = number of medical responses from i to s,
X^2_{is} = number of mechanical responses from i to s, and
T^* = minimum emergency medical response time established for each freeway section or a standard throughout.

All other terms are the same as those used by the authors.

It is fully recognized that the model suggested may not be applicable to practical use for determining actual freeway service station locations. However, the authors are urged to explore the possible expansion of their model to include some of the above characteristics. With the tremendous emphasis on the attainment of high air quality standards by the mid-1970s, the authors are urged to modify their model to include, as a secondary output, measures of air pollution for alternative service facility systems.

In conclusion, the authors are to be congratulated for an excellent exploratory modeling effort that appears to have great potential for practical application.

Joseph A. Wattleworth, University of Florida

The authors have presented a very important step toward the development of an analytical tool to assist in the selection of an optimal system of facilities to provide emergency service to freeway motorists. The purpose of the service facilities is to respond to incidents, which reduce the freeway capacity, and to act to restore the capacity of the freeway as quickly as possible. The trade-offs involved in such an analysis are (a) the cost of the service facilities provided versus (b) the reduction in delay cost to the freeway motorists.

The analytical technique described is a combination of the freeway simulation model, which was previously developed by May and others, and a model of another type. The latter model can be one of a number of kinds, such as integer programming or dynamic programming models. The simulation model is used to determine the relationships between service level, accident location, and service location. In this way it is possible to determine which service locations would provide an adequate service level for any freeway section. When the feasible service station locations are determined for each freeway section, the second model is used to determine the optimal service station locations. This is done essentially by minimizing the number of these service stations.

Any mathematical model makes certain assumptions in describing a complex real-world situation. These assumptions are necessary to accomplish the abstraction desired. They must, however, be borne in mind when the results of the model are interpreted, when the use of the model is considered, or when further developmental work on the model is considered. Some of the assumptions made by the authors will be discussed in the following paragraphs. The importance of several of these assumptions is recognized by the authors, and they have suggested them as further research.

The first two assumptions are made in the simulation model. These are that there is only one accident per peak period and that there is a constant capacity loss for all accidents. These suggest an addition to the simulation process to add a Monte Carlo generation of the occurrence of accidents and the severity of the accidents. These data would then be input into the existing simulation program (with some modifications).

The model as currently formulated considers only one direction of flow. To be practical, of course, the model will have to optimize service facility locations with regard to both directions of the freeway. It would appear to be a rather straightforward extension of the model to make this change.

As formulated, the model uses maximal values of delay as a basis of the optimization. It is more traditional to use expected values for this purpose so that one can compare the annual cost and annual benefits of all candidate systems. The expected values of the delay time due to accidents could be obtained by use of the Monte Carlo accident generation routine with the existing models.

The assumptions examined so far can be relaxed by extensions or modifications to the existing models. These are some assumptions that do not appear to be so easily handled within the framework of the existing models. One of these is the assumption of stationary service facilities, e.g., garages that house service vehicles waiting to be dispatched to an accident scene. An alternate approach, such as is practiced on the freeway system in Chicago, is that of emergency patrol vehicles. These vehicles are not housed in fixed-location facilities but rather move in the traffic stream. They sometimes arrive at accidents that are unreported and are sometimes dispatched to the accident scene. In any case, it is doubtful whether this type of system could be considered by the reported models or their extensions.

The model also apparently assumes that the same service facilities are provided at each service station. This will probably not be the case in a real-world application. An examination of the case in which several accidents can occur in one peak period and the inclusion of less service incidents, such as disabled vehicles, will probably lead to the need for several patrol vehicles per service station. If the number of vehicles per service station is not constant, the cost of each service station will not be constant. If the variability of cost is quite high, the optimization routine will have to be changed because it minimizes the total cost by minimizing the number of service stations. If the cost per station varies, the total system cost cannot be determined by multiplying the number of units in the system by the unit cost.

The discussions have centered on several of the assumptions that were made in the reported models. As such, the discussions should not be construed to convey a negative evaluation of the paper or the models. The authors are to be congratulated for undertaking the analysis of the problem of optimizing freeway service facilities. This is an excellent example of a real-world problem being submitted to analysis rather than a mathematical model in search of an application. More such applications of analytical techniques to real problems are needed. The authors did not solve all of the problems associated with the optimal design of freeway service facilities but have made an important step toward such an optimization.

AUTHORS' CLOSURE

The authors are appreciative of the thoughtful and valuable reviews by Hess, Carter, and Wattleworth. The three discussions are mainly concerned with the simulation model (FREEQ) and the assumptions that were made throughout the paper. The writers agree that the initial assumptions are limiting and that further investigations are definitely desired in order to have more valid results.

As Carter pointed out, short time fluctuations in traffic demand may occur. One should carefully select the appropriate length of time slice such that the relative fluctuation is small.

In the paper, it has been mentioned that there are two types of problems: minimal total delay time and minimal response time. It then seems reasonable that the first type of problem is relevant to police or mechanical service, whereas the second type of problem is more important to ambulance or fire squad service. If these four services are considered to be independent of each other, the problems can be treated separately. (This has been illustrated in Figure 1, the schematic model.) Of course if one would not think in this way then two sets of constraints, one for the total delay and the other for the response time, might be included in one single problem.

The freeway capacity of each subsection is estimated, based on the method described in the Highway Capacity Manual (3). The capacity

$$C = 2{,}000 \times W \times N \times T$$

where

W = width factor,
N = number of lanes, and
T = truck and grade effect.

Detail of this was given elsewhere (4, pp. 29-32). The travel time of service vehicle in each time slice and subsection, on the other hand, was obtained directly from the simulation result.

The multi-incident case is much more complicated. One has to know the probability distribution of the incidents over the time-distance space first. There are a number of papers that discuss the probability model from a macroscopic aspect. However, it is more desirable to find a model that can be used to predict the secondary incident. On the other hand, the estimated service time of a dispatched service vehicle (including response time, on-site service time, and travel time to return to the station) must be found. Perhaps, for this problem the best one can do is to simulate the actual performance of the service vehicles under different systems and traffic conditions. Wattleworth suggested the use of the Monte Carlo approach to solve the problem. Usually simulation is a good method that can be used to analyze a complicated system whenever any analytical way seems helpless. A lot of effort, however, may be required to do this. If there exist 10 candidate locations, for instance, then a total of $2^{10} - 1$ feasible solutions must be considered, one for each simulation run. It then appears that the central problem for simulation technique is how to reduce the computation effort.

The reported models are established to deal with stationary service only. Certainly, it cannot be employed to analyze patrolling service systems. Simulation results (1, 2) based on the data from the San Francisco-Oakland Bay Bridge showed that stationary service systems were better than patrolling service systems because of higher benefit-cost ratio. However, this may not be the general case. An interesting study could be made by comparing these two systems under different traffic conditions.

The optimization model is essentially to search for all the minimal set solutions. Different minimal set solutions may have a different number of elements (number of locations). If the incurred cost of each location is known, it is simple to compute the total cost for each minimal set of solutions. If there is a cost limitation, some of the solutions may be eliminated in the course of dynamic programming computation. Furthermore, different types of service facilities are not necessarily located at the same place. As was mentioned earlier, if each of the basic services is treated independently, then, for each type of service, one has to solve the problem once.

Hess made a critical comment about the assumptions used in the paper regarding capacity reduction due to incidents. Under normal conditions (in case of no incident), most of the freeway users, if necessary, would make a weaving movement far ahead of the place where the number of lanes is changed, provided that the passing motorists are familiar with the physical configuration of the freeway. On the other hand, if an incident occurs and blocks some of the lanes, a drastic reduction in the capacity will happen prior to the incident location, because of the weaving effect. Texas Transportation Institute made an interesting study that showed that, when an accident occurred in a three-lane freeway, the average flow was reduced by about 50 percent if one lane was blocked and 70 percent if two lanes were lost. This clearly demonstrated that the relation between reduction in capacities and lanes is not proportional. Certainly, more study on this is needed.

The speed-flow relationship suggested in the Highway Capacity Manual is not always applicable to a particular study area. The car-following theory has been studied for many years; there are a number of papers that investigate the relationship among flow, density, and speed. Unfortunately, none of them can be used as a general model. Probably, it is desirable to annex several subprograms in FREEQ, each corresponding to a specified car-following model, and to leave it as an option for the program users.

It is fully understood that the present report did not cover all the problems concerned with the optimal design of freeway service facilities. There exist a number of versions of the problem. This paper is only one phase of a broad research area. Some of the assumptions in the paper were used only to simplify the computation effort. Much detailed study and further work are needed.