Simulation and Analysis of Helicopter Emergency Transport in Upstate New York

ORIE 4580 Final Project

Annie Cheng (zc375), Yixin Huang (yh454), Yingxi Li (yl877), Joe Ye (ty357), Hanyu Zhang (hz248)

Executive Summary

Helicopters serve as important means of transportation for emergent medical services due to their immediacy and flexibility. Therefore, deciding where and how to base the helicopters in order to boost efficiency and provide better services is a crucial task for the helicopter dispatch center. In this project, we aim to provide a specific set of recommendations on how many helicopters should be employed and where they should be based in upstate New York so as to minimize the response time of the helicopters to the emergent calls, boost the fraction of calls that get responded, and make the best use of each employed helicopter to avoid unnecessary cost. We tackle this task by constructing a discrete-event simulation model that replicates the helicopter transportation process. By running a search-selection method on the model, analyzing the changes in performance measures, and manually experimenting with different arrangements based on the observations, we present two sets of helicopter configuration recommendations (for every fixed number of helicopters from 1 to 12): one solely prioritizing the average response time for calls, and the other allowing some trade-offs between the average response time and the response fraction. To further assist the decision-making process, for each one of the 24 candidate configurations, we holistically take into account its performance measures, the number of helicopters used, the number of base locations used, etc., and arrive at two final recommendations:

1) **{Buffalo: 2, Rochester: 1, Ithaca: 1, Syracuse: 3, Albany: 1}** with 8 helicopters, 29.24 minutes of average response time, 58.7% of response fraction, 3.146 of helicopter utilization;

2) **{Buffalo: 2, Rochester: 3, Syracuse: 5, Albany: 2}** with 12 helicopters, 33.54 minutes of average response time , 72.7% of response fraction, 2.600 of helicopter utilization.

The two configurations have their own distinct advantages: configuration 1) prioritizes the minimization of response time, and with only 8 helicopters, it significantly boosts the helicopter utilization. This configuration should be used when a very quick response is desired over all other factors or when clients only want to use a limited number of helicopters. Configuration 2) has a high response fraction while pertaining to a reasonably low response time in the trade-off. It uses only 4 bases, which also saves the fixed cost. It should be used when a balance of high response fraction and low response time is needed. We remain optimistic that a number of other similar configurations yield slightly different but equivalently competent performances, as shown in our full result table, and could be adopted based on client needs. The following sections of the report provide detailed descriptions of the problem, our modeling approach and implementations, our selection methods and results, and analysis and discussion of the modeling process.

I. Problem Description

Helicopter transport has become a crucial part of the medical care system as it efficiently transports patients with urgent medical needs from their injury scene to medical facilities. However, helicopter transports are extremely expensive, and should only be used when the patient's need for instant medical care is sufficiently high. Therefore, when a call is made by a patient, a first responder will arrive at the scene, provide high-quality first aid, and assess whether helicopter transport is needed. If such demand is deemed necessary, the call will be transferred to the helicopter dispatch (HD), delivering information on the number of patients requiring transport and their locations. HD decides whether the current weather is safe to fly and dispatches the closest available helicopter to the scene. If not, the call will be recorded as "not serviced - unsafe". The call will also be recorded as "no helis" if no helicopter is available within the range. For the scope of our project, there are only 5 helicopter bases available to be chosen from Buffalo, Rochester, Elmira, Ithaca, Sayre PA, Watertown, Syracuse, Binghamton, Utica, and Albany due to a huge amount of fixed costs required to set up a helicopter base.

Calls might be canceled during the process: if the call is canceled before the helicopter arrives at the scene, possibly due to the death of the patient, the helicopter will return to its base; the call cannot be canceled after the helicopter has reached the scene. Once the helicopter arrives at the scene, it will transport the patient to a medical facility (hospital/trauma center) and then return to the base.

The goal of this project is to consider how many of the 12 helicopters should be employed and where they (and their crews) should be based around the upstate New York region to effectively respond to emergent calls. We want to optimize the following performance measures: percentage of calls dispatched, average response time, time to definitive care, response fraction, and utilization of helicopters.

II. Modeling Approach and Assumption

We use discrete-event simulation to model the problem described above. Specifically, our model contains the following sequence of events:

a. Call arrives: Calls arrive at the HD and check if there is any available helicopter.
b. Calls canceled: Cancel the current call and the associated helicopter returns to base.
c. Calls finished at HD center: Check if it is safe to fly. If so, the helicopter completes preparation and departs for the scene; if not, the call is canceled.
d. Helicopter departs for calls: Call might be canceled before the helicopter arrives at the scene.
e. Helicopter arrives at scene: Helicopter arrives and stays at the scene for a period of time.
f. Helicopter departs from the scene: Helicopter transports the patient to a trauma center or a hospital based on the proximity and urgency of the patient.

g. Helicopter arrives at hospital: Helicopter arrives at the hospital and spends some time until departure.

h. Helicopter departs hospital: Helicopter has finished serving for the call and returns to the base.

i. Helicopter arrives at base: Helicopter arrives at the base.

The connection between events is summarized in Figure 1.

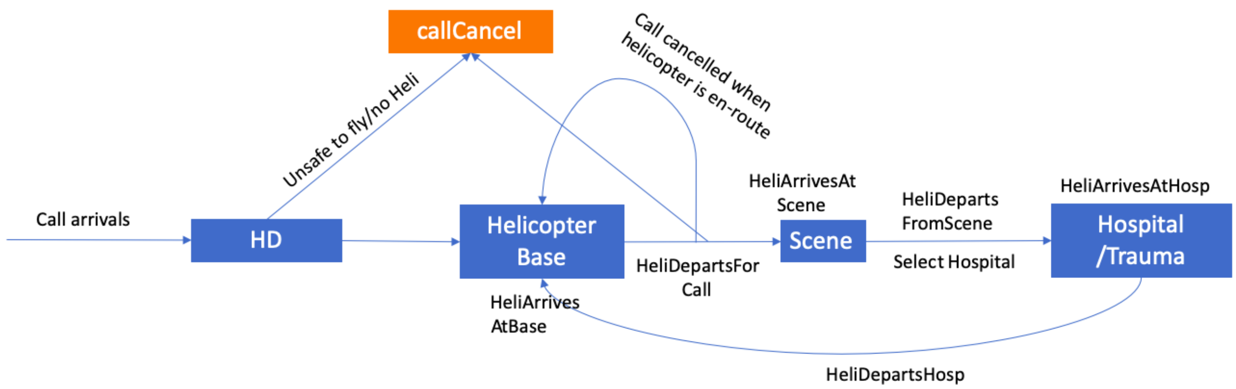
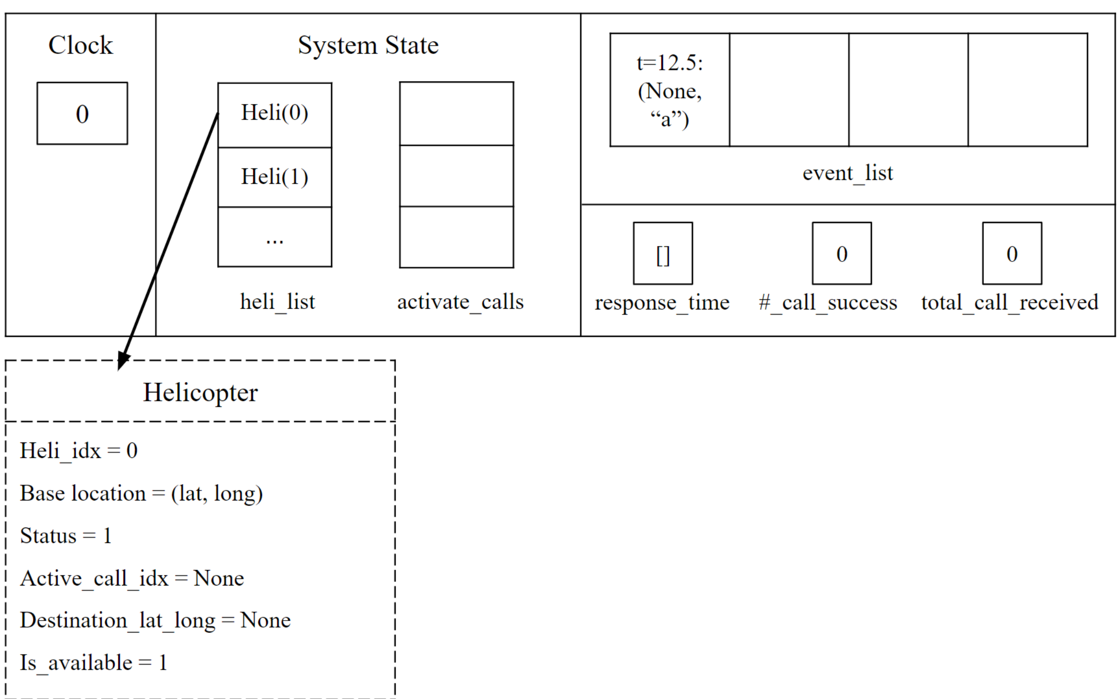
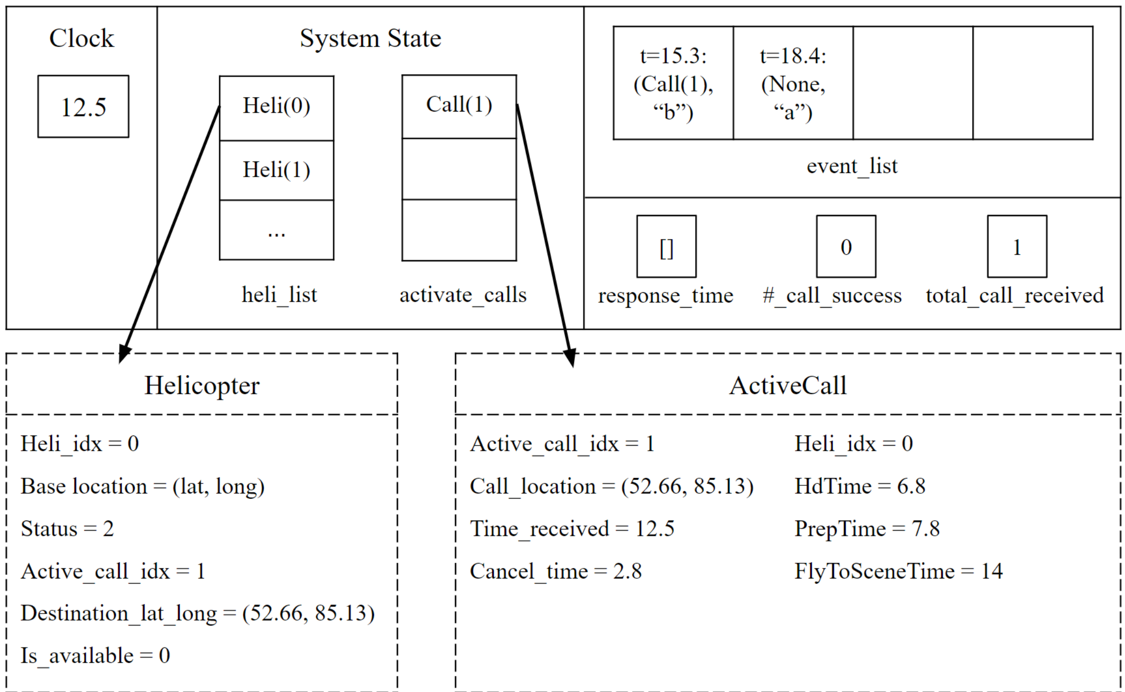


Figure 1. Flow Chart of the Discrete-Event Simulation

To better understand the events in the model, we create the following snapshots of our discrete-event simulation model at time 0 and at one succeeding event time.



(a)



(b)

Figure 2. Simulation System Events

Figure 2a shows an instance of the system, including states, events, and statistical counters when we first launch the simulation. The variables in the *Model* class are shown inside the concrete box at the top, while the variables in the *Helicopter* and the *ActiveCall* classes are shown inside the dashed box at the bottom. We start with clock = 0, with an empty active_calls list, and a heli_list containing default *Helicopter* objects which are all at the base. In addition, we generate the first event in the event_list: "a" denotes a call arrival event; t = 12.5 is the generated time when the first call arrives; None means we have not created an ActiveCall object that links to this event. Moreover, the three measures are listed, including response_time, #_call_success, and total_call_received, which are necessary for the calculation of performance measures.

Figure 2b shows the succeeding event of 2a. When clock = 12.5, which is the time of the first call arrival event, we update the system. Firstly, the total_call_received is updated to 1. We then create an *ActiveCall* class object, stored as Call(1) in the active_calls list. We generate its call_location, assign it to the nearest available helicopter with index heli_idx, and generate the HdTime, PrepTime, and FlyToSceneTime from respective distributions. In the meantime, the status of the call's corresponding helicopter is updated to 2, meaning that it is now waiting for clearance, thus changing is_available to 0. Since we've completed the call arrival event at t = 12.5, we remove it from event_list and generate a new call arrival event. In this instance, the second event we generate for Call(1) is "b", denoting cancellation of the call, because its cancel_time is less than HdTime.

This instance is merely a snapshot of our system. Other events update the system in similar ways. More technical details of the model can be referred to in the appendix. We believe this discrete-event simulation approach works efficiently in the problem setting and provides us with ample information for making recommendations.

Finally, we make the following assumptions as we build the model:

a. The helicopters have a traveling speed of 160 km/hr and can respond to calls within 180 km of their base. They always have sufficient fuel and servicing between calls.
b. In reality, helicopter crews have limited hours in which they can fly in a single shift. The HD staff might also be reluctant to dispatch a helicopter crew when they are close to the end-of-hour shift. However, we are not considering these complications in our model. Instead, we allow the helicopters to be dispatched at any time.
c. Helicopters in our model must return to base before they can be redirected to a new call, however in reality they can go directly to the new scene.
d. In reality, many other factors may go into whether it is safe to fly, such as the weather and the route, however, we assume our decision is purely random with a certain probability without regard to these considerations.
e. We also assume that each call transports only one patient to a single location, without considering the possibility of multiple patients' transportations.

A most detailed implementation of our model can be found in Appendix 1.

III. Data Analysis

We are given a dataset containing records of all calls received by the HD over a year. Details for each record include the call location, the hospital location, the indicators of helicopter availability ("Heli Avail") and fly safety ("Safe to fly"). Based on the data given, we generate the input distributions of the model:

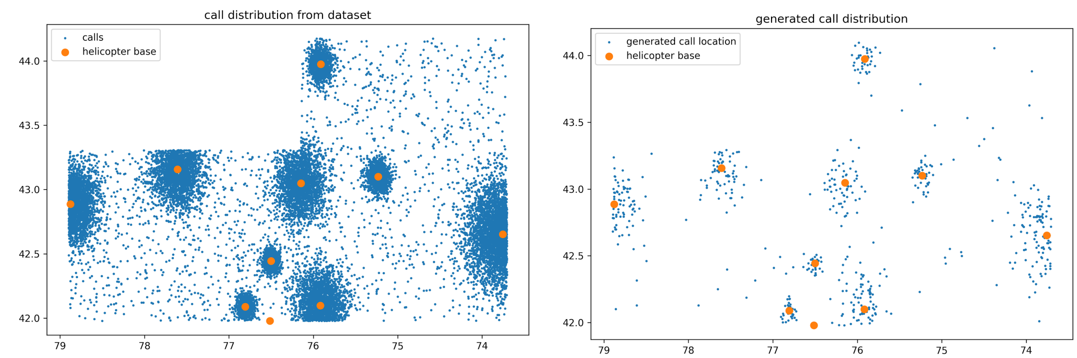
a. Call Arrival Rate: We fit a nonstationary Poisson process to the Time column in the dataset. We assume that the rate is constant for an hour and compute the rate of calls for each hour in a day. See the exact rates of each hour in Appendix 6.
b. Call Locations: To determine the call locations, we divide the horizontal and vertical sides of upstate New York each into 100 equal segments; thus, in total, we have $100 \times 100 = 10{,}000$ squares. We calculate the probability of calls in each square, then we assume that calls are uniformly distributed within the square. See the distribution in Appendix 7.

c. Safety to Fly: There is a possibility that the call cannot be handled due to unsafe conditions. We generate the probability using Bernoulli Distribution with a success rate of .899. The success rate is implied by averaging over the “Safe to fly” column in the dataset.
d. Helicopter Base Locations: Helicopter bases are being considered in Buffalo, Rochester, Elmira, Ithaca, Sayre PA, Watertown, Syracuse, Binghamton, Utica, and Albany. We are limited to choosing only 5 bases from this list of locations to base up to 12 helicopters in total. We use the longitude and latitude of each location in the simulation model.
e. Helicopter Delay Time at Dispatch Office: We use Triangular Distribution to generate the delay time at the helicopter dispatch office. We are given that the min and max to be 5 and 10 respectively, and the mode to be 7.
f. Helicopter Delay Time for Flight Check and Preparation: Again, we use Triangular Distribution with min and max to be 5 and 10, while the mode is equal to the average of min and max, which is 7.5.
g. Cancellation Time Distribution: In the model, we generate a cancellation time for every flight if it is safe to fly. However, the cancellation only occurs if it happens before the helicopter arrives at the scene. We generate the cancellation time as an Exponential Distribution. To determine the parameters of the distribution, we refer back to our dataset. The Cancel Delay field in the dataset gives us the time after the call was first received to when it was canceled. We then use Maximum Likelihood Estimation to calculate the parameter for the Exponential Distribution and obtain $\lambda= 0.205$.
h. Scene Time: To determine the time spent at the scene, we fit and overlay various distributions to the data and determine which is a good fit. We fit Gamma, Beta, Lognormal, Exponential, Pareto, and Weibull, and find that Beta(a=2.95, b=11072, scale=1302) and Gamma(a=2.95, scale=0.12) are both good fits. We use the Beta distribution in our model.
i. Hospital Time: Similarly, we fit various distributions to the dataset and find both Beta (a = 2.90, b = 812, scale = 141) and Gamma(a=2.91, scale=0.17) fit well. Again, we choose to use the Beta distribution in the model.
j. Hospital Locations: There are 4 trauma centers, Rochester, Syracuse, Albany, and Sayre PA, as well as 6 other medical facilities. Calls whose nearest hospital is a trauma center will be sent to its nearest automatically; among the remaining, 0.807 fractions of calls in the dataset are still sent to their nearest medical facility (non-trauma-center), while the rest are redirected to a trauma center that is farther. We will use the fraction of 0.807 in our model to mimic this process.

IV. Model Verification

After building the model, it is important to verify that it works properly.

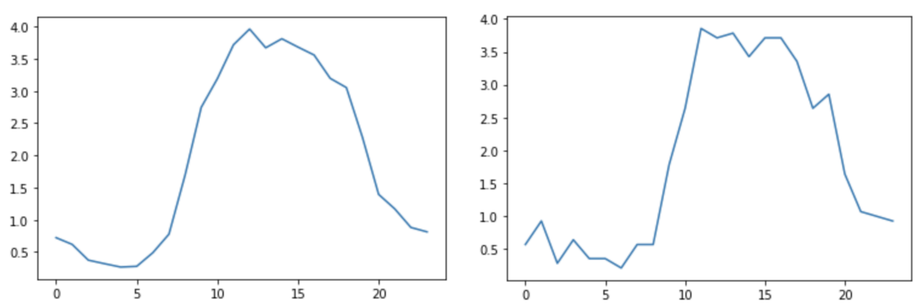
a. Call locations generation: Figure 3 displays two scatter plots: the first plot is the scatter plot of the call distribution from the dataset and the second one is the generated call distribution. We notice that the two scatter plots are highly correlated in terms of the locations of the clusters as well as the size and density of each cluster.



a) Call locations in the dataset b) Call locations generated in our model

Figure 3. Call location distribution

b. Call arrival generations: to ensure that the call-arrival rate works as intended, we plotted the average number of calls per day in a two-week simulation period (as shown in Figure 4 right). This corresponds to the nonstationary Poisson rates (i.e. the arrival rates) from the dataset (as shown in Figure 4 left). We see that they follow a similar distribution, which ensures that our model is generating calls at a correct rate as intended.



a) Poisson rate from the dataset b) Average number of generated calls

Figure 4. Call arrival rate distributions

c. Trace: Figure 5 shows a snippet of the helicopters' status tracing in a replication. It is used for verifying the correctness of the model by ensuring that helicopters go through stages in an anticipated way with sensible time elapses. For example, if we trace helicopter 6 in Figure 5, we can see that at time 19536, it departs from the scene, it then arrives at the hospital at time 19540. Afterward, it departs the hospital and comes back to base with reasonable time elapses, which further verifies the correctness of the model. We can see that a new call is assigned to helicopter 6 at the time 19602.

```
Helicopter 6 departs from scene at time 19536.424586439585
Helicopter 2 arrives at hospital at time 19539.869468432025
Helicopter 6 arrives at hospital at time 19540.031362921472
Helicopter 3 departs hospital at time 19541.43145135273
Helicopter 3 is back to base at time 19541.44145135273
Helicopter 5 assigned to safe HD call at time 19547.91568929231
Helicopter 1 departs from scene at time 19549.640787994653
Helicopter 6 departs hospital at time 19550.553165446618
Helicopter 6 is back to base at time 19550.563165446616
Helicopter 5 departs for call at time 19556.127072486153
Helicopter 4 departs from scene at time 19556.52806005363
Helicopter 4 arrives at hospital at time 19557.510151705632
Helicopter 1 arrives at hospital at time 19559.859553337432
Helicopter 5 arrives at scene at time 19561.788485633
Helicopter 9 arrives at scene at time 19566.61440545465
Helicopter 7 assigned to safe HD call at time 19568.857710443568
```

```
Helicopter 2 departs hospital at time 19569.55208109041
Helicopter 2 is back to base at time 19569.562081090407
Helicopter 4 departs hospital at time 19573.957625470073
Helicopter 0 assigned to safe HD call at time 19583.021549032488
Helicopter 3 assigned to safe HD call at time 19585.687070067845
Helicopter 9 departs from scene at time 19588.041080090643
Helicopter 0 departs for call at time 19591.50606438786
Helicopter 5 departs from scene at time 19594.78498493647
Helicopter 9 arrives at hospital at time 19597.94217460129
Helicopter 1 departs hospital at time 19598.180722345285
Helicopter 5 arrives at hospital at time 19600.446398083313
Helicopter 6 assigned to safe HD call at time 19602.3195483152
Helicopter 6 departs for call at time 19607.941671891855
Helicopter 6 is cancelled on the way back at time 19614.114959326
787
Helicopter 4 is back to base at time 19618.76700917095
```

Figure 5. Helicopter Traces

V. Performance Measures

There are 5 performing measures we are focusing on in this project, which are the percentage of calls dispatched, response time, time to definitive care, response fraction, and utilization of helicopters. We define a call to be successful if the patient is picked up by a helicopter at the scene.

a. Percentage of calls dispatched = $\frac{\textit{Calls which have helicopters dispatched}}{\textit{All calls}}$

b. Response time = *Time(helicopter arrives at scene) - Time(a call received at HD)*

c. Time to definitive case = *Time(patient reaches receiving facility) - Time(a call received at HD)*

d. Response fraction = $\frac{\textit{Number of successful calls}}{\textit{Total number of calls}}$

e. Utilization of helicopters = $\frac{\textit{Average number of successful calls per day}}{\textit{Number of helicopters}}$

When we select our optimal configurations, we mainly consider 3 of the 5 performing measures, which are the utilization of helicopters, average response time, and response fraction. Percentage of calls dispatched is neglected since

$$\frac{\textit{Calls which have helicopters dispatched}}{\textit{All calls}} = \frac{\textit{Successful calls} + \textit{Calls cancelled halfway}}{\textit{All calls}},$$

which shows that the percentage of calls is highly correlated with the response fraction of the calls. Time to definitive care is another factor we do not consider further in model selection since

$$\textit{Time}(\textit{patient reaches receiving facility}) = \textit{Time}(\textit{helicopter arrives scene}) + \textit{Time}(\textit{from scene to receiving facility})$$

Our model has no control over the time spent from the scene to the receiving facility because it is constant once the call location and hospital location are fixed. Therefore, the time to definitive care can be reflected from the average response time.

VI. Model Selection

The main goal of the project is to find the best sets of configurations of helicopter base locations that optimize the performance measures described above. Thus, the decision variable of our model is the base location of each helicopter, along with the number of helicopters.

First of all, we draw inspiration from the greedy algorithm to generate a complete configuration (12 helicopters) that minimizes the average response time. This means, at the moment, average response time is the only performance measure being taken into consideration. Starting with only one helicopter, we loop over all helicopter bases to decide which base optimizes its average response time. This requires us to check 10 cases. We then fix the base location of this first helicopter and introduce the second to create a configuration of two helicopters. Again, we loop over all 10 bases to decide where to base the second helicopter -- here, average response time is still our sole measurement. So on and so forth, we consider 10 extra cases when a new helicopter is introduced to the configuration at each step. As soon as 5 base locations have appeared in the current configuration, we restrict the rest of the steps to search within only the five bases that have already been used. We continue the greedy searching process until a complete configuration of 12 helicopters is found, and claim this configuration gives the best locally optimal average response time of 27.4 minutes per call. Similarly, we find another complete configuration that achieves the best possible response fraction of 73.9%. Note that the complete configuration that maximizes response fraction also has the greatest utilization because, with the number of helicopters fixed at each step, a higher response fraction implies more successful trips and thus a greater number of successful calls handled by each helicopter.

An advantage to this approach is that the complete configuration is constructed in orders of helicopter “importance”: in each step, one more helicopter is added to the base that maximizes the marginal benefit. For example, the first helicopter base location must be a locally optimal step that contributes the most to the overall performance measure, whereas the second helicopter location, although also critical, might not yield as much of an improvement as the first one because it is not chosen in the first step. The earlier a placement of a helicopter (represented by a helicopter base index) appears in the complete configuration, the more contribution it has to the overall performance.

At this point, we have two arrays of length 12 sorted by importance to the performance measure of the fraction of response or average time of response. Utilization will only be considered later when comparing configurations with different numbers of helicopters.

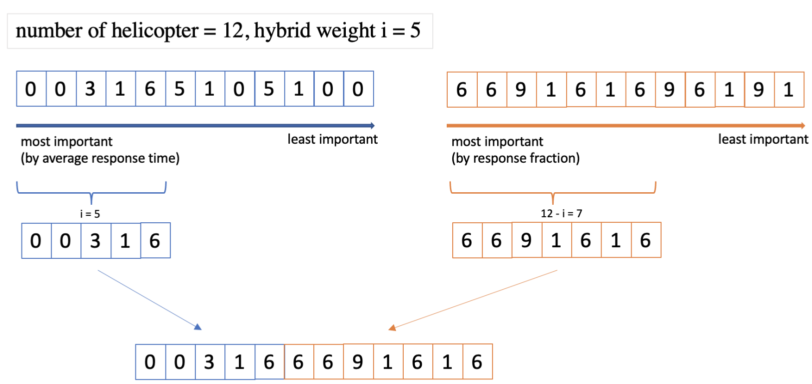


Figure 6. An example illustrating our result hybrid process (n=12, i=5)

To generate a result that both minimize the average response time and maximizes the fraction of response, we hybridize the two ordered lists of helicopters as follows: first, we iterate over the possible number of helicopters using integer n from 1 to 12; then, we iterate over a hybrid weight integers i from 0 to n. For each fixed i and n, we construct a new configuration with the first i locations from the complete configuration corresponding to response time and the first n-i locations from the complete configuration for response fraction. Figure 6 illustrates the hybridization process for $n=12$ helicopters when hybrid weight $i=5$. We take the first 5 base locations in the average response time result and the first 7 base locations in the response fraction to get a new solution of 12 helicopters. For each n from 1 to 12 (number of helicopters), we plotted all response measures for each hybrid configuration against their hybrid weight (i on the x-axis), as shown in Figure 7. Hybridizing and plotting allow us to catch some great configurations that perform adequately well on all performance measures. For example, if we look at the plot with 5 helicopters (row 2, column 2), configuration #1 clearly outperforms its neighbors with a high response fraction and a locally minimal response time.

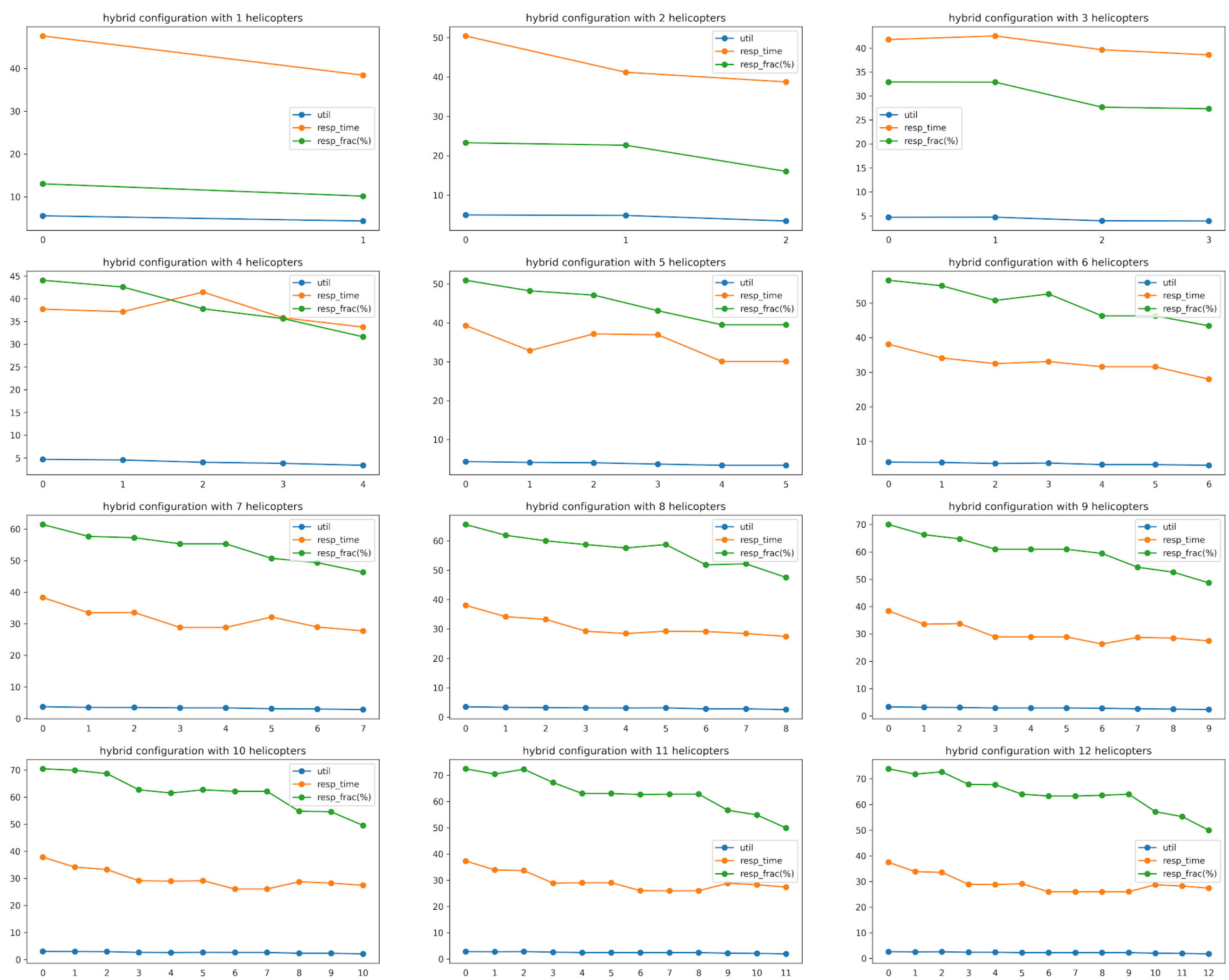


Figure 7. Hybrid configuration performances

To verify the selection method we used above and gain more insight into the results we've obtained, we adopt the Random search method. We do not expect Random search to be the most optimal method for this problem since the dimension of the decision variables (base locations for helicopters) is not fixed, which will result in high runtime. As a result, we only use a budget of 1000 (10 replications for each simulation) to run the Random search method for each of the performance measures. The results (see Figure 8 below) are consistent with those of the selection method mentioned above. Solely considering one performance measure, the maximized utilization of helicopters obtained by Random search is 5.44, which is obtained by having one helicopter in Utica. The maximized response fraction and minimized response time obtained using Random search are 59.8% and 28.4 minutes, respectively. Random search returns worse results than our selection method, possibly due to the low budget used. However, Figure 8 demonstrates that with an increased number of incumbent best solutions being searched, the curves of all three performance measures converge to the optimal objectives obtained by our selection method. Therefore, we are confident with proceeding with our selection method.

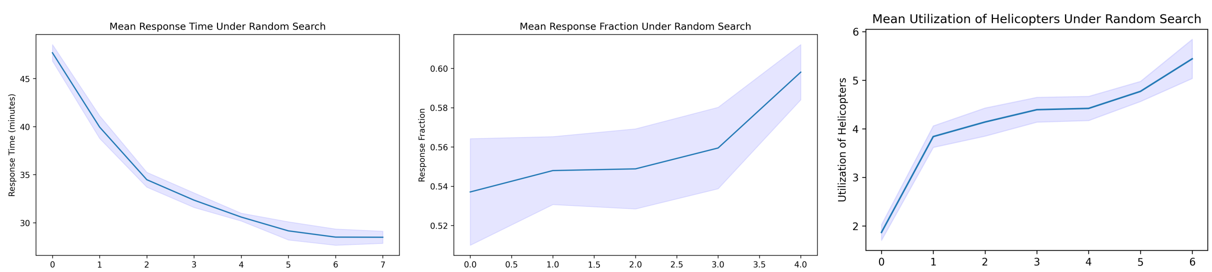


Figure 8. Mean performance measures with 95% confidence interval using Random search

VII. Results

We can discover a clear trade-off between the response fraction and the average response time, which is exactly what we expected. For each number of helicopters (n from 1 to 12), we scrutinize the visualization in Figure 7 and select 2 configurations: one basic recommendation that primarily minimizes the average response time, and one alternative recommendation that leans moderately towards maximizing the response fraction while still guaranteeing a reasonably small average response time. For some fixed number of helicopters, one configuration might be used as both the basic recommendation and the alternative recommendation, so long as it outperforms other configurations by demonstrating both a lower response time and a higher response fraction. Table 1 and Table 2 correspond to the basic recommendation set and the alternative recommendation set respectively, where the detailed helicopter configuration, and the associated average response time, response fraction, helicopter utilization are listed.

Num_Heli	Bufflo	Rochester	Ithaca	Syracuse	Albany	resp_t(min)	resp_frac	util
12	2	3	1	4	2	28.789	0.677	2.419
11	2	2	1	4	2	28.9	0.672	2.619
10	2	2	1	4	1	29.137	0.627	2.689
9	2	2	1	3	1	28.906	0.61	2.903
8	2	1	1	3	1	29.238	0.587	3.146
7	2	1	1	2	1	28.851	0.553	3.387
6	2	0	1	2	1	33.085	0.526	3.756
5	1	1	0	2	1	32.861	0.482	4.132
4	0	1	0	2	1	37.735	0.441	4.702
3	0	0	0	2	1	41.793	0.329	4.683
2	1	0	0	1	0	41.186	0.227	4.855
1	1	0	0	0	0	38.449	0.102	4.357

Table 1. Basic recommendations (prioritizing average response time)

Num_Heli	Buffalo	Rochester	Syracuse	Albany	resp_t(min)	resp_frac	util
12	2	3	5	2	33.542	0.727	2.598
11	2	2	5	2	33.674	0.722	2.816
10	2	2	4	2	33.289	0.687	2.944
9	1	2	4	2	33.565	0.663	3.158
8	1	2	4	1	34.177	0.619	3.315
7	1	2	3	1	33.519	0.577	3.53
6	1	1	3	1	34.115	0.55	3.927
5	1	1	2	1	32.861	0.482	4.132
4	0	1	2	1	37.735	0.441	4.702
3	0	0	2	1	41.793	0.329	4.683
2	1	0	1	0	41.186	0.227	4.855
1	1	0	0	0	38.449	0.102	4.357

Table 2. Alternative recommendations (mildly sacrificing average response time for higher response fraction)

We provide our clients the option to choose between the two sets of configurations according to their preferences and specific needs. If a small average response time needs to be achieved at all costs, we would recommend the basic set of solutions; if the average response time can be mildly sacrificed to yield a more favorable response fraction, we would recommend the alternative set of recommendations. Moreover, if our client cares most about a high helicopter utilization, then we recommend them to pick from the lower end of the recommendation table, where a small number of helicopters are used.

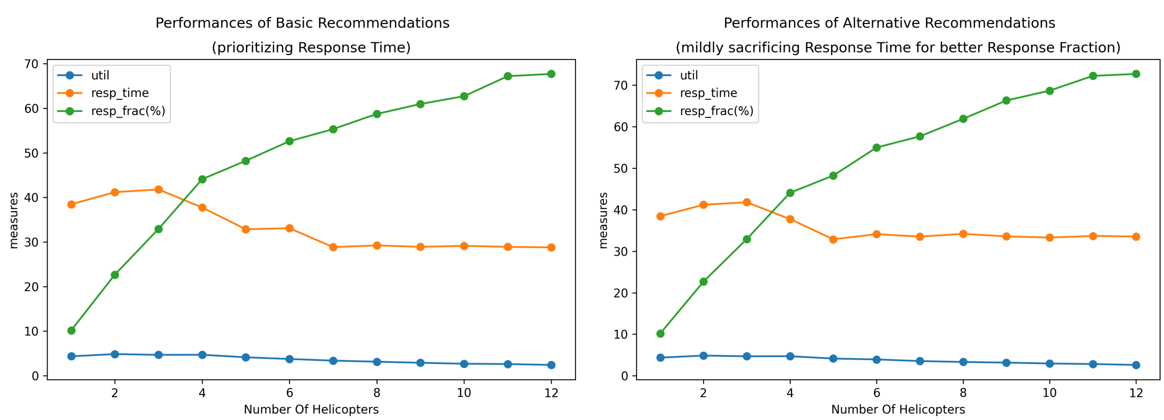


Figure 10. Performance measures of all 24 recommendations

To facilitate comparison and interpretation, we plot the performance measures in each set of recommendations against the number of helicopters (see Figure 10). As the number of helicopters increase, we expect lower response time, greater coverage and response fraction, as well as a slight decrease in helicopter utilization. The major difference across the two plots is that the basic recommendations end up converging to a response time that is lower than the alternative recommendation. Yet, the alternative recommendations demonstrate a greater potential in achieving a higher response fraction (above 70%) and meeting more emergent medical service demands.

From the above results, we make the following two recommendations:

1. {Buffalo: 2, Rochester: 1, Ithaca: 1, Syracuse: 3, Albany: 1}. This recommendation optimizes the average response time to 28.85 minutes while only using 8 helicopters, at the expense of a 58.7% response fraction. It should be used if a low average response time is desired.
2. {Buffalo: 2, Rochester: 3, Syracuse: 5, Albany: 2}. This recommendation improves the response fraction to 72.7%, at the expense of increasing the average response time to 33.54 minutes, using all 12 helicopters. It should be used if a better response fraction is desired.

VIII. Sensitivity Analysis

We further investigate the robustness of the model by conducting sensitivity analysis on the following parameters: safe-to-fly ratio, percentage to the nearest hospitals, helicopter speed, and call arrival rate. Specifically, we look into whether the changes in the parameters have impacts on the three key parameters: the average response time, the response fraction, and the utilization of helicopters. For safety to fly ratio and percentage to the nearest hospitals, we look into the differences in performance measures when changing 5% and 10% of the original percentage employed in the simulation model. As for helicopter speeds, we run simulations from 140 km/h to 180 km/h. In addition, we perform sensitivity analysis on the call arrival rates by comparing the situations of 50%, 75%, 150%, and 200% of the original model's arrival rate

We first consider the safe-to-fly ratio. This parameter can easily vary in real life since it not only depends on the weather condition but also the path a helicopter takes from its base to the scene. As can be seen from Figure 11, when the safe-to-fly ratio changes, the 95% confidence intervals of the average response time are highly overlapped. This indicates that changes in safe to fly ratio have no significant impact on the average response time. This makes sense because response times are only recorded if the helicopters are safe to fly. Therefore, the average response time is unaffected by the safe to fly ratio as expected. As for the fraction of response (Figure 11 right) and the utilization of helicopters (Appendix 2), they increase as the safe-to-fly ratio increases. This is especially evident in the case of the fraction of response since the confidence intervals of different safe-to-fly ratios do not overlap. This is expected to happen. Since increasing the safe-to-fly ratio would lead to the increasing number of helicopters departing from HD, the number of successful calls is accordingly increased. As the number of successful calls serves as the common numerator of the fraction of response and the utilization of helicopters, the latter two performance measures are expected to increase.

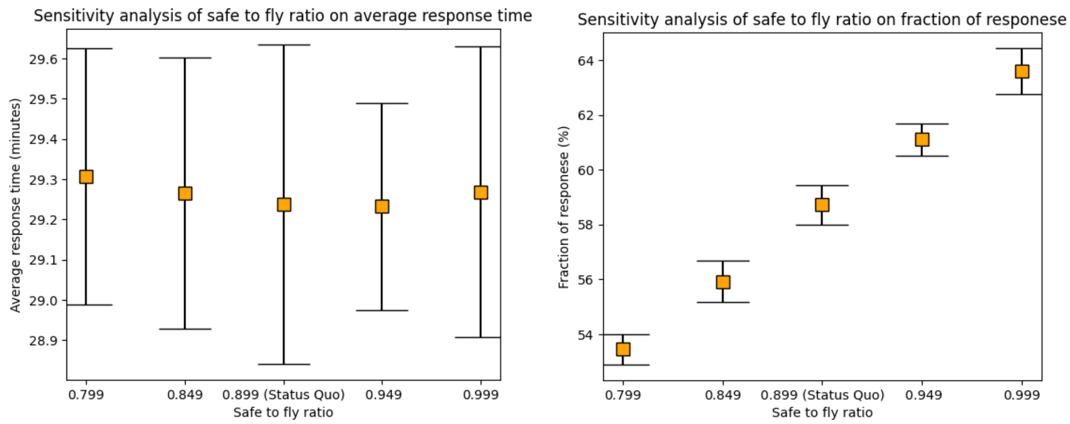


Figure 11. Sensitivity analysis of the safe-to-fly ratio

The percentage to the nearest hospital is a parameter of interest because this ratio might easily vary due to patients' health conditions. In the season of pandemics, this ratio might increase as patients would need to go to the trauma center more frequently instead of the nearest hospital. As for the effect of changing the percentage to the nearest hospitals, we expect the utilization and fraction of response to slightly increase and the average response time will be unaffected. This is because when the percentage to the nearest hospital increases, the time spent for each helicopter in the flight from the scene to the medical facility is smaller; therefore, the number of successful calls can increase, so as the utilization of helicopters and response fraction will also increase. Nonetheless, the graphs in Appendix 3 show that both the utilization of helicopters and the response fraction only exhibit minor changes and most of the confidence intervals overlap. Although the results of sensitivity analysis slightly contradict our expectation, this is understandable since the change in the percentage to nearest hospitals might only have minor impacts on the total number of successful calls. The effect could potentially scale up if the percentage to the nearest hospitals changes dramatically, which seems unlikely in reality.

Helicopter speed can vary for different types of helicopters and different weather conditions. Our sensitivity analysis results (see Appendix 4) show that changes in helicopter speed have the most significant impact on the average response time as the confidence intervals of different helicopter speeds barely overlap. This decreasing trend is in tune with our intuition: for the same distance needed for travel, faster helicopter speed would result in a smaller amount of time required for transportation. While there's an increasing trend for both response fraction and utilization of helicopters, the changes are minor as the confidence intervals have huge overlaps with each other.

The last parameter of interest is the call arrival rate, which, again, might be heavily dependent on the season. The sensitivity analysis results (see Appendix 5) indicate that while the call arrival rate has almost no influence on the average response time as the confidence intervals highly overlap, the response fraction and the utilization of helicopters change drastically as the call arrival rate increases. When applying this simulation model to real life, we should take further care of the call arrival rate as it might potentially have a huge impact on the performance measures.

IX. Conclusions

In this report, we attempt to streamline helicopter transport in situations of medical emergencies in upstate New York by proposing a set of recommendations on where to place the helicopters and how many helicopters to place at each base. We identify three primary responses to evaluate the performance of each potential recommendation, including helicopter utilization, average response time, and response fraction, which turn out to have trade-off relationships: utilization tends to drop whereas the other two responses tend to improve as more helicopters are used; response fraction and average response time tend to result in very different base choices. Therefore, we propose two full sets of recommendations, one leaning towards lower average response time at the expense of a lower response fraction, and another aiming for a higher response fraction while guaranteeing a relatively low average response time. Each set contains twelve different recommendations from using one helicopter to using twelve helicopters, such that clients could use as many helicopters as they possibly can. From these two recommendation tables, we took into consideration all factors and picked 2 recommended configurations to present in order to provide more assistance to the decision-makers.

Our model still contains some limitations. In the model, we simplify the helicopter emergency transport process as we ignore hangover shifts, the time limit a helicopter crew could fly, and the possibility of a helicopter flying directly from the hospital to the next call locations. In our single-response-based selection, we only choose the local maximum at each step, which means we might potentially ignore some better configurations that require the search of higher dimensions. Future exploration of this problem may include incorporating more detailed aspects in the model. For example, future researchers can take into consideration helicopter shifts or use other search-selection methods to look for results with better performance measures.

Appendices

Appendix 1: Implementation of the Model

We use Python 3.9 to build the discrete-event simulation through object-oriented programming. In the "ORIE4580_Simulation_Project" folder, we have the following files:

File	**Description**
main.py	This module contains the *Model* class
heli_call_class.py	*Helicopter* and *ActiveCall* class for the model
test.py	Test cases to verify our model
selection.py	Functions to select base locations and helicopter placements
sensitivity.py	Sensitivity analysis on different parameters to validate the model
randomsearch.py	Function to implement Random search

Our Python modules are made up of three classes: the *Helicopter* class, the *ActiveCall* class, and the *Model* class. The class *Helicopter* has these variables: heli_idx, base location, status (numbers 1 to 8 denote whether the helicopter is at base or OnWayToScene or somewhere in between), active_call_idx (None if at base), and destination_lat_long (None if at base). In addition to the getters and setters, the class also includes a function *is_available* that checks whether the current helicopter is available for a new call.

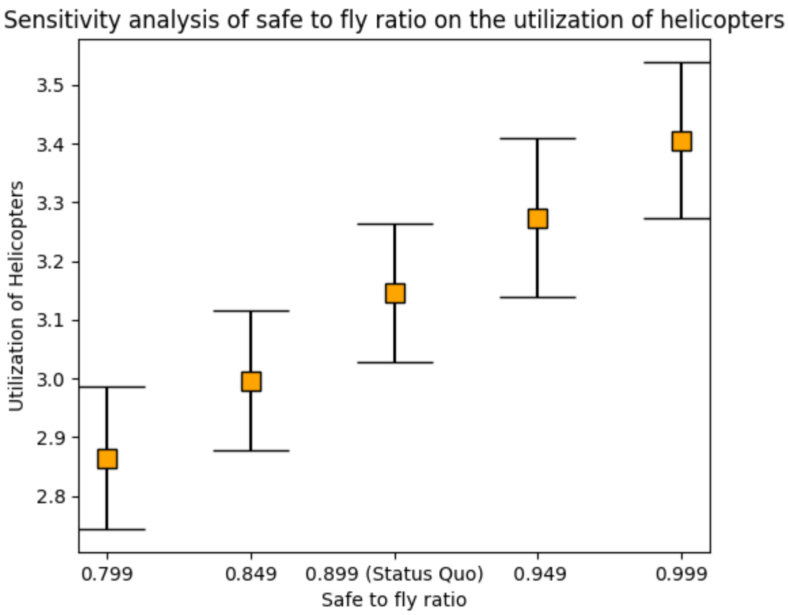
The class *ActiveCall* represents an active call object. The class has these variables: active_call_idx, call_location, time_received, cancel_time, Heli (that is of type *Helicopter*), hdTime, prepTime, and flyToSceneTime.

The class *Model* creates instances of simulation. The default setting is with a warm-up period of one week and a run length of four weeks (including the warm-up period). The variables are stored in a dictionary named *fixed_factors*, taking input from users when constructing the class object. Another dictionary *specification* stores the fixed parameters. We define the metric to keep track of in a dictionary named *metric* and the system states of the model in another dictionary named *states*. To initialize, we create helicopter objects and store them in a list; start the system clock at 0; create an empty list of active calls; create an empty event list; create a list of random number generators with a fixed stream and different substreams. In addition, the following 10 helper functions are defined to help represent each of the events as illustrated in the modeling approach section.

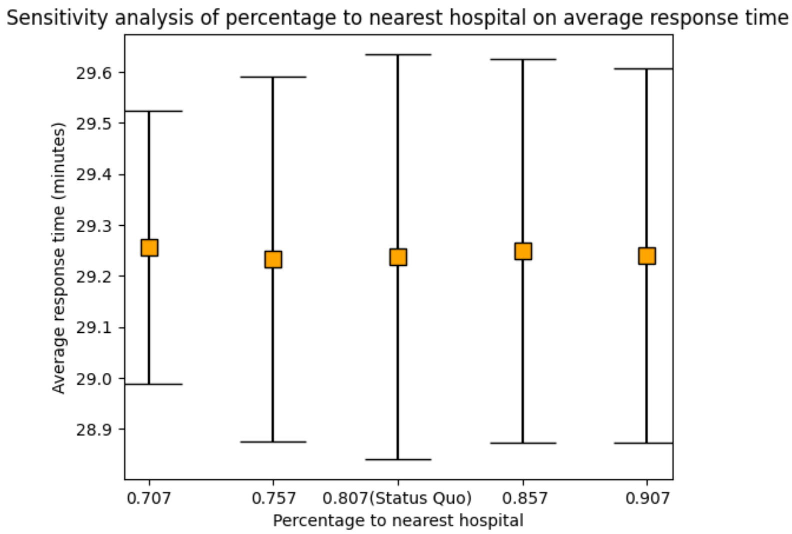
Helper Function	Description
CallArrival	Generate a CancelTime for the new call and check if there is any available helicopter; if there is, create a call object and push it to the active call list. Generate the next CallArrival event.
CallCancel	Remove the call from the active call list and adjust the helicopter status.
CallFinishesAtHD	Generate random outcomes safe to fly and update the helicopter status accordingly. Generate the time at which the helicopter completes prep and schedule CallCanceled/ HeliDepartsForCall event.
HeliDepartsForCall	Update the status of the helicopter to be OnWayToScene. Compare the call cancellation time and the arrival time of the helicopter at the scene, then adjust the helicopter destination accordingly.
HeliArrivesAtScene	Generate the time helicopter at the scene, update the helicopter status and add the event HeliDepartsFromScene.
HeliDepartsFromScene	Send the patient to a Trauma Center or Hospital based on the Bernoulli r.v. and adjust the helicopter destination accordingly. Update the helicopter status; compute the time the helicopter arrives at the scene, and add the event HeliArrivesAtHosp at that time.
Select_hospital	Select the nearest Hospital/Trauma Center when the helicopter departs from the scene.
HeliArrivesAtHosp	Update the helicopter status to AtHosp. Add an event HeliDepartsHosp at the current time plus a sampled hospital time.
HeliDepartsHosp	Remove the call from active calls and update the helicopter status to OnWayToBase. Compute the arrival time of the helicopter at its base and add the event HeliArrivesAtBase at that time.
HeliArrivesAtBase	Update the status of the helicopter to “AtBase”.

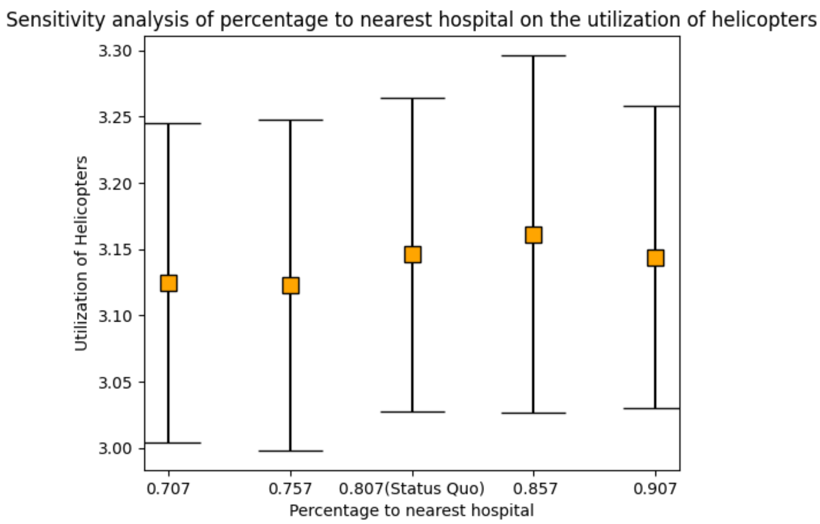
Moreover, to reduce variance and obtain comparable outputs when simulating different systems, we employed the common random number method. Specifically, the RNG package given in class is applied. For each model, we use the same stream but apply different substreams for different random distributions. Besides, the subsubstreams are advanced for each replication to ensure independence.

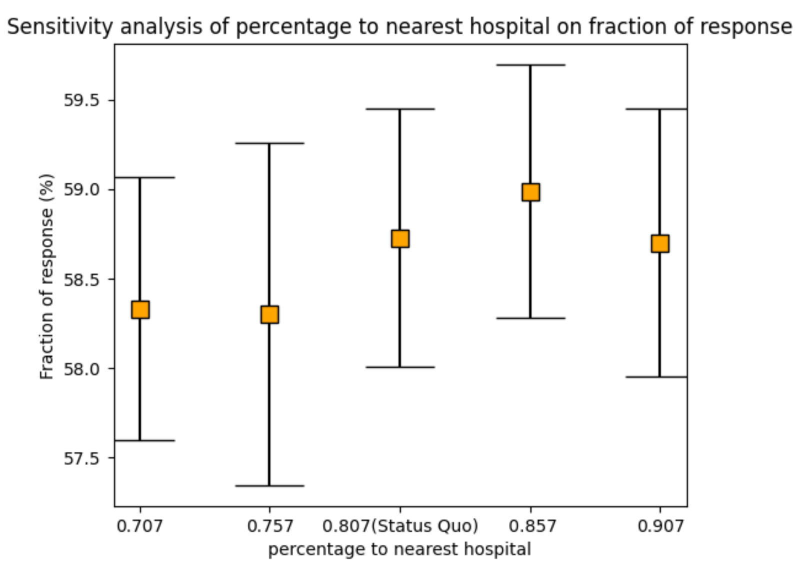
Appendix 2: Sensitivity analysis of safe-to-fly ratio on the utilization of helicopters



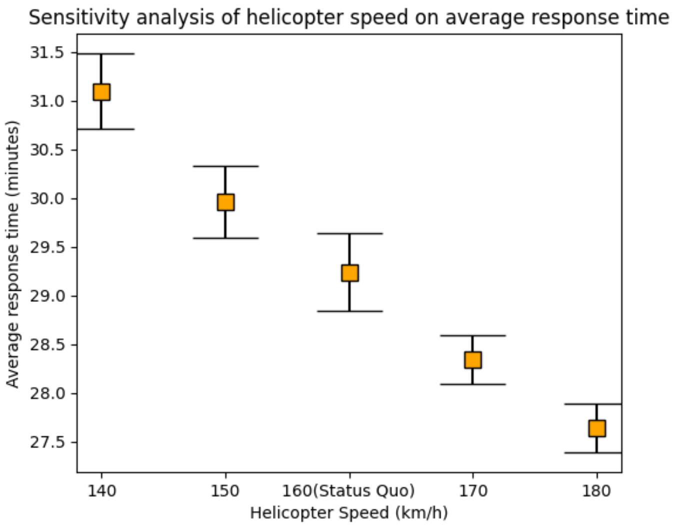
Appendix 3: Sensitivity analysis of percentage to nearest hospital

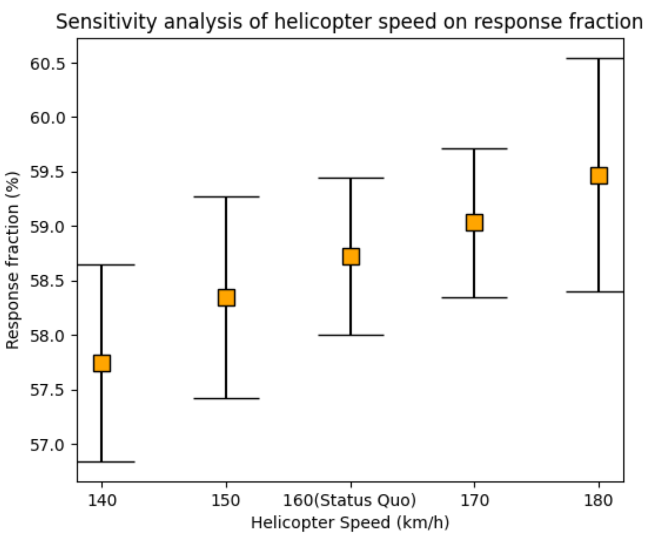


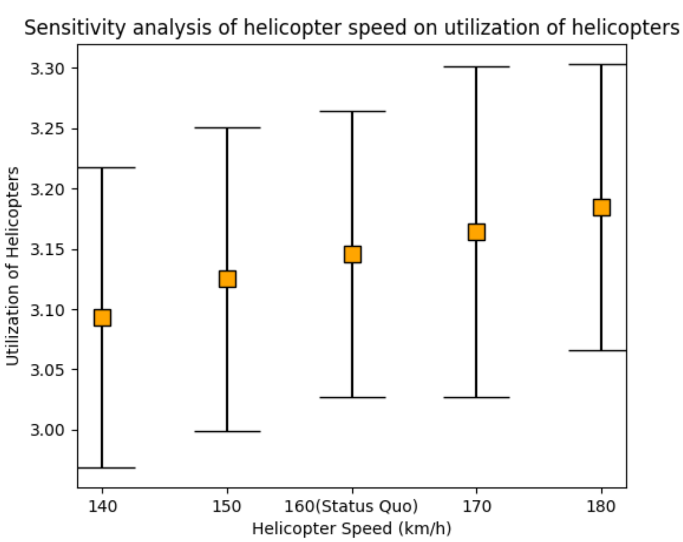




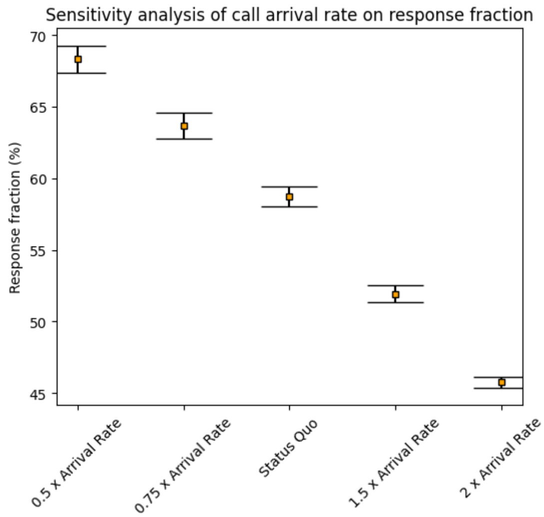
Appendix 4: Sensitivity analysis of Helicopter Speed

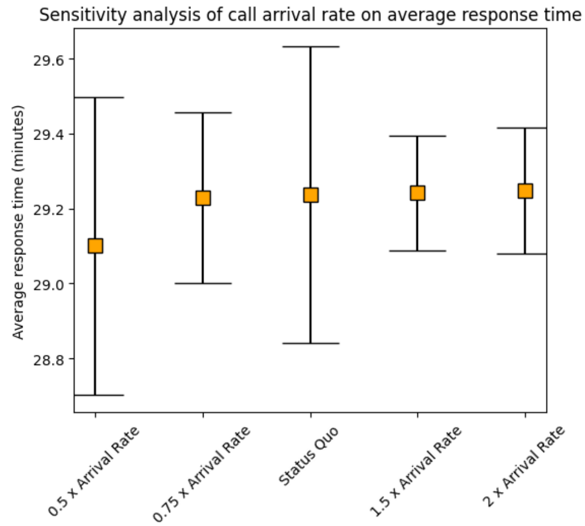


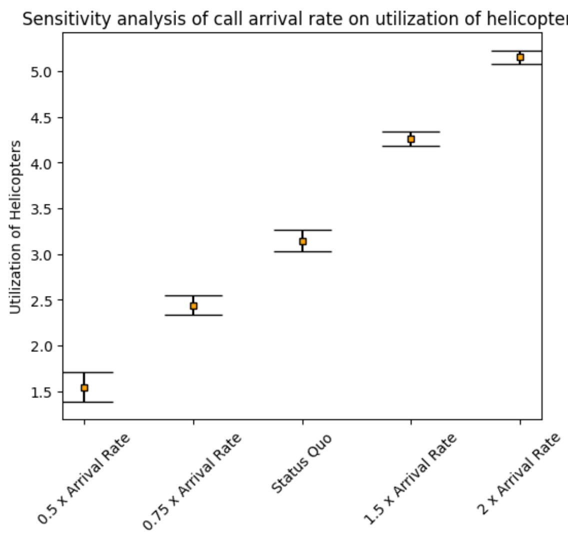




Appendix 5: Sensitivity analysis of call arrival rate







Appendix 6: Call arrival rate distribution

Hour	rate
0	0.722527
1	0.618132
2	0.373626
3	0.318681
4	0.266484
5	0.277473
6	0.489011
7	0.777473
8	1.695055
9	2.747253
10	3.192308
11	3.719780
12	3.964286
13	3.673077
14	3.813187
15	3.684066
16	3.560440
17	3.197802
18	3.054945
19	2.280220
20	1.395604
21	1.170330
22	0.881868
23	0.813187

Appendix 7: Call Location Density