Helicopter Emergency Transport in Upstate New York

Jianyang Zhou, Kevin Jiang, Ashley Jiang, Cher Xu, Qiyu Yang

Executive Summary

Helicopters are used in air medical services to transport urgent/emergent patients, and it is very costly to use helicopter transportation. We are tasked with finding the optimal locations for helicopter bases, given a number of helicopters between 1-12, to maximize the number of calls dispatched and minimize response time. We will use important performance measures such as the percentage of calls dispatched, average response time, and utilization of helicopters to determine where these optimal locations are.

To model our inputs, we are able to fit common distributions to our data and incorporate randomness in our inputs. For the model itself, we simulate a sequence of random events which include flight preparation, flight cancellation, time spent at scene, etc. using our inputs. We are able to run replications of this event list on a randomly generated list of calls.

We discovered that the best option is to spread 9-10 helicopters around 6-7 base locations. The best locations for the helicopter bases are in Ithaca, Binghamton, Buffalo, Rochester, Syracuse, Watertown, Elmira. Since there are more helicopters than optimal base locations, Buffalo, Rochester, and Watertown are good locations to support more than one helicopter if 9-10 helicopters are deployed. We used an iterative process to determine these optimal locations, firstly finding an ideal location for 5 helicopters and using intuition, call locations, and other heuristics to place extra helicopters. For example, more helicopters are placed at hospitals with a higher relative density of calls surrounding its area.

Although it seems intuitive that using all 12 helicopters would yield the best performance, this is not actually the case. We find that having 12 helicopters is actually excessive and wasteful because helicopter utilization begins to decrease when there are too many. Response time does decrease, but not significantly, and there is a large trade-off between this slightly better response time and high costs.

Ultimately, the heuristic proves acceptable, seeing as optimizing locations for 6 and 7 bases using brute force yields the same result as our heuristic.

Problem Description

The goal of this project is to determine the optimal locations for emergency helicopter bases in the upstate New York region, with a given number of helicopters between 1-12. In this region, there are 10 hospitals, 4 of which acting as trauma centers for patients requiring specialized treatment. 5 of these hospitals will be

base locations. We will also explore the possibility of having more than 5 bases and examine if this possibility will yield faster response times to calls/fewer calls that could not be responded to.

One of the main issues in determining where these bases should be placed is managing uncertainty in the travel of helicopters. The determination of whether it is safe to fly, the flight preparation time, whether the flight will be canceled or not, and the scene and hospital time are all random variables, and we will determine which random variables should represent each of these times. Another main issue is simulating call locations, which are also randomly spread across the upstate New York area.

We are interested in several performance measures, primarily the percentage of calls dispatched, average response time, and utilization of helicopters. Percentage of calls dispatched and response fraction are closely related, as well as average response time and time to definitive care, so we are mostly considering the three aforementioned performance measures. The performance measures help us determine the optimality of the model with different base locations. We will examine how small changes to the model change these performance measures through sensitivity analysis.

Modeling Approach & Assumptions
Modeling Approach

To achieve the ultimate goal of optimizing the aforementioned performance metrics, our model situates a given number of helicopters at fixed, known locations and simultaneously tracks and handles the sequence of events generated by every helicopter upon receiving an (emergency) call. In particular, a general description of a complete sequence of events for a single helicopter responding to a single call is as follows:

(i) A call is made from a particular location in the upstate New York area, which is then received by the helicopter dispatch (HD) station. (ii) After spending some time making a decision about whether the call is safe to respond to or not (e.g. a call may be deemed unsafe due to the presence of inclement weather), the nearest available helicopter will be dispatched to the call's scene location, if the call is deemed safe. (iii) During the time it takes to take off and reach the scene
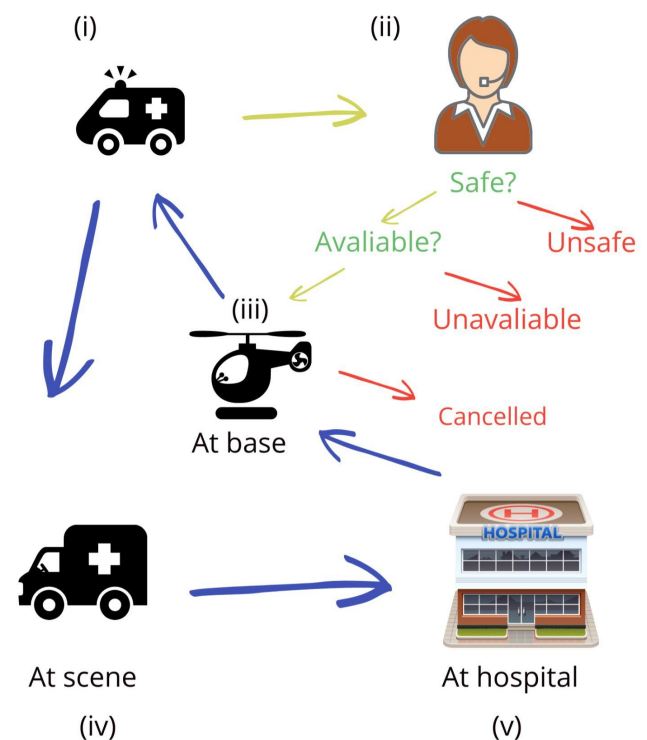
location, the call may be canceled due to the patient's death, a medic later deeming transport unnecessary, etc. This possibility is considered in our model implementation. (iv) Upon arriving at the scene location, the helicopter will transport the patient to the nearest hospital or, if the patient requires special treatment, the nearest major trauma center. (v) Finally, after unloading the patient, the helicopter travels back to its base and waits until HD dispatches them to another call.

As in nearly every complex problem, the project's underlying scenario begets a large number of variability throughout the development of a single call: for instance, there is inherent randomness and uncertainty in the location at where and the rate at which calls emerge, how long it takes to make a decision at HD, how long it takes to travel between locations, etc. If not explicitly addressed in the assumptions section below, we model any sources of randomness with an appropriate statistical process and/or probabilistic distribution that accurately reflects the information obtained from previous data concerning helicopter response times in emergency calls in the upstate New York area. We ask the interested reader to please refer to our *Data Analysis* and *Appendix* sections below.

Assumptions

In developing the model, we make several key assumptions on various entities in the simulation model. For ease of comprehension, we have split our assumptions into a few categories based on the entity in question.

Helicopter Behavior

We assume the following concerning the operational limitations and behavior of the helicopters: (i) helicopters do NOT require refueling or servicing; (ii) they have no restrictions in how many hours they can fly in a single shift (i.e. the workers remain the same throughout the entire simulation); (iii) they can be dispatched at any time including graveyard hours; (iv) they cannot be redirected mid-flight; and, (v) they travel at a fixed constant speed of 160 km/hr and can respond to calls within 180 km of their base.

Patient/Call Behavior

We assume the following concerning each call and its associated patient: (i) call requires the transport of only one patient, which gives rise to the unique pairing of a single call and patient; (ii) the safeness of responding to a call is determined without regard to the base or scene location; (iii) the call may be canceled in between the time the HD receives the call and the time a helicopter reaches the scene of the call; and, (iv) the call is never canceled (i.e. the patient does not die) after the helicopter arrives at the scene, since the patient is there receiving intensive care.

General/Miscellaneous Assumptions

Lastly, we take the liberty of making the following assumptions concerning the overarching architecture of our problem statement and model:

1. We ignore day-of-week and seasonality effects and assume that there are only time-of-day effects.
2. We assume that hospitals have unlimited capacity and that no hospital would turn away a patient that a helicopter brings to it.
3. We are assuming that the helicopter bases should be located in the cities in which hospitals/trauma centers are located. This seems more logical than placing helipads sporadically around upstate New York, even if this yields more optimal results. Thus, **we impose the condition that the emergency medical staff work at these hospitals**, and therefore helipads should be built at hospitals so they can easily board these emergency helicopters.

Data Analysis

As alluded to in the *Model Approach & Assumptions* section, we need to model and fit probabilistic distributions to necessary input parameters to our simulation. Using these distributions for certain events permits us to incorporate randomness into the data inputs and, consequently, accounting for the inherent variability present in the real-world scenarios associated with our project problem.

Call Location Generation

In order to simulate call locations, we utilized the given project data to give the likelihood of a call falling into a certain area. To generate the locations of calls, the map of upstate New York is gridded into 35 by 35 bins, where the probability of a call landing in each bin is estimated. Then, within each bin, the location is assumed to be uniformly distributed inside the bin. Using this method, we are able to generate a heatmap of upstate New York that is very similar to the heatmap of the given project data but incorporates the randomness of the calls.
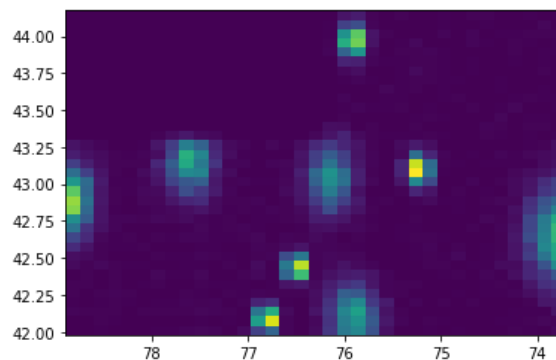

Figure 1: Simulated call locations using gridding method where brighter colors indicate higher call density

Flight Safety Determination and Flight Preparation

The time required at the helicopter dispatch to receive the details of the call and make an assessment on flight safety is usually about 7 minutes but can be anywhere from 5 to 10 minutes, as given. Therefore, we may use the Triangular distribution with parameters (5, 7, 10) to model this delay. As for the flight preparation, we are using a Triangular distribution with parameters (5, 7.5, 10), as all we are told about the flight preparation time is that it "takes between 5 and 10 minutes."

Flight Cancellation

In order to determine if the flight is canceled, we use an exponential distribution with a rate of 0.205 hr to determine cancellation time or CT. This is given to us as to how cancellation time will be modeled. If CT is below the time it takes to take off, then the flight is canceled and the helicopter remains at its base. If CT is greater than 30 minutes, then it is censored. If CT is greater than take-off time, 30 minutes or less, and less than the time to reach the scene, the helicopter stops en-route to the call scene and returns to its base. The final case handled is when a CT is 30 minutes or less, but the helicopter is already on the way back from the call site, then it is treated similarly as if there was no cancellation and the helicopter returns to its base with the patient, presumably deceased.

Time Spent at Scene

Using the project data, we are able to fit a gamma distribution with parameters (2.95, 0.12) to the Scene time. We use this to represent the time spent on the scene of the call. During this time spent on the scene, there is a determination made about whether the patient needs specialized treatment, and in this instance, they will be sent to a trauma center. This may be the closest hospital to them, and notably, in the project data, each time a patient is not brought to the hospital closest to them, it's because that hospital was not a trauma center and they needed specialized treatment. The probability that they are sent to the hospital closest to them from the project data is 0.807. As a result, the probability that we use in our model to represent a patient needing specialized treatment at a hospital that is not the closest to them is 0.193.
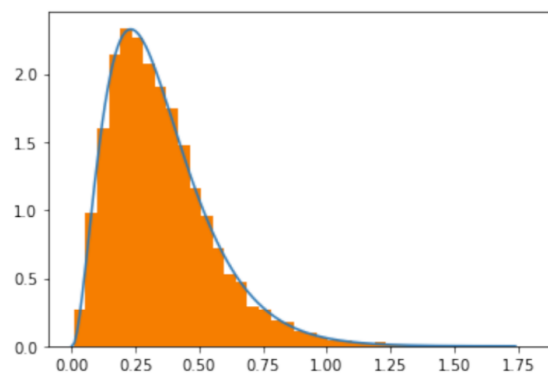

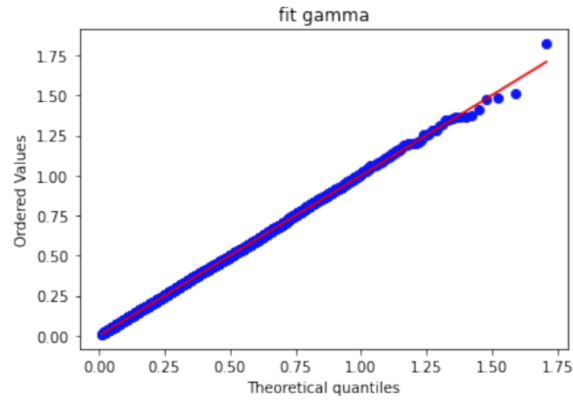Figure 2: Fitting the gamma distribution to the scene time

Figure 3: Q-Q plot of gamma distribution for scene time

Accounting for Hospital Times

In the project data file, we are given data for "Hosp time," which represents the amount of time the helicopter spends at the hospital. This Hosp time fits a gamma distribution with parameters (2.91, 0.17), so we use this to determine the time each helicopter spends at the hospital.
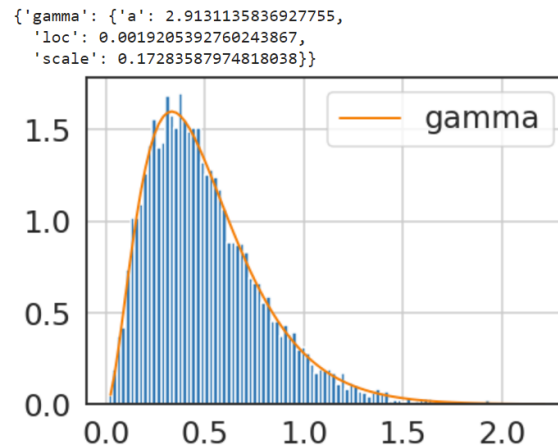

Figure 4: Fitting the gamma distribution to the hospital times

Model Verification

It is natural to ask whether or not our implementation of the model described above with its nuanced interplay of various events properly works. To these ends, we have verified the model's validity via (i) a path-following event sequence check, and (ii) a theoretical queuing test based on the model's results/performance metrics. We now describe these as follows:

(i) The path-following event sequence check consists of taking the sequence of events associated with each individual entity (i.e. calls or helicopters) and tracking the entity's progression until its termination. In broad strokes, we follow the entity until its termination, which may be one of several possibilities (e.g. a call may be deemed unsafe, unavailable, or successfully responded and transported), and then compare our observed number of each type of entity with the number we expect. The latter number can be computed without any simulation since it is based solely on given and/or estimated probabilities; for instance, the probabilities of whether or not a call is deemed safe to fly to, or, more challenging but doable, whether or not a call can be responded to with an available helicopter. We found that our numbers aligned remarkably well, which indicates that calls and helicopters logically progress throughout the simulation in our desired manner.

(ii) The theoretical queuing test is applied to each hospital center by treating a hospital as a simplified queuing system from which we can obtain precise theoretical results and then compare against our empirical results from the simulation model. In particular, we consider each hospital to be a server system with a limited queue capacity wherein each 'server' can be thought of as a helicopter and the 'customers' as emergency calls (to the more simulation savvy, we consider an M/M/c/N queuing model where 'N' is precisely 0, since either a call can be responded to or not). From queuing theory, we obtain exact results concerning our desired performance metrics such as the fraction of customers/calls responded to and server/helicopter utilization. However, we found that our theoretical results slightly overestimated the fraction of calls responded to and the utilization of the helicopters, indicating that the helicopters were deemed to be more efficient than they actually were in the simulation. We hypothesize the discrepancy arises from (a) having to make several rough estimations of key parameters in the simplified queuing model: in particular, the overall arrival rate of the calls within a 180 km radius (namely, the area to which helicopters can respond to) and the average service time for each hospital was obtained via a simple average or multiplication over input model distributions; (b) hospitals may overlap in various parts of the upstate New York area, in which case, fewer helicopters would need to be employed to meet the same fraction of emergency calls. Regardless of this slight difference between the theoretical and observed performance metrics, we believe our model works as intended and can be deployed to run the necessary experiments to tackle the project problem. We do so in the following section *Model Analysis & Sensitivity Analysis.*
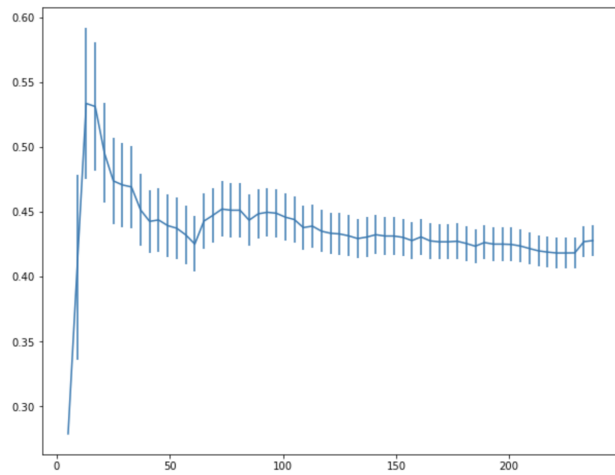
Model Analysis & Sensitivity Analysis

Model Analysis

Figure 5: Determination of warm-up period

Before we start answering any question with simulation, we want to first set a reasonable warm-up period by plotting the expectation of one performance measure and observe when it plateaus. We chose the average response time and we got the following plot by calculating the expected response time for every four hours. And by observing this graph, we set the warm-up length to be 72 hours.

To find the model that minimizes the average response time, the direct way is to enumerate all possible combinations of helicopter's location from 1 to 12 helicopter situations. If we place them in 10 bases, the total combinations of helicopter's location are:

n	1	2	3	4	5	6	7	8	9	10	11	12
combin-ations	10	55	220	715	2002	5005	11440	24310	48620	92378	167960	293930

If we place them in 5 bases, the total combinations of helicopter's location is:

n	1	2	3	4	5	6	7	8	9	10	11	12
combin-ations	10	55	220	715	2002	4795	10060	19015	33130	54127	83980	124915

By repeating the simulation experiment and taking the average, we got the optimized solution for situation n = 1 to n = 5(95% confidence interval),

#Helicopters	Based	Calls Dispatched	Avg. Response Time(hours)	Responded	Utilization of Helicopter
1	Albany	0.120	0.631	0.109	10.82

		(0.1192, 0.1217)	(0.6244, 0.6379)	(0.1078, 0.1103)	(10.77, 10.87)
2	Binghamton, Albany	0.281 (0.2794, 0.2830)	0.603 (0.5992, 0.6063)	0.254 (0.2524, 0.2558)	10.82 (0.1077, 0.1087)
3	Watertown, Binghamton, Albany	0.401 (0.3989, 0.4024)	0.593 (0.5901, 0.5961)	0.362 (0.3602, 0.3636)	12.04 (12.01, 12.07)
4	Watertown, Binghamton(2), Albany	0.488 (0.4859, 0.4907)	0.584 (0.5814, 0.5866)	0.441 (0.4386, 0.4431)	11.00 (10.0097, 11.0003)
5	Ithaca, Binghamton, Buffalo, Rochester, Syracuse	0.575 (0.5726, 0.5777)	0.567 (0.5648, 0.5695)	0.519 (0.5166, 0.5212)	10.82 (10.77, 10.87)

These results answer the questions of where we should place our helicopters when we are limited to 5 bases only and have 5 or fewer helicopters. The placement of 5 helicopters (Ithaca, Binghamton, Buffalo, Rochester, Syracuse) is especially important because they can potentially serve as a default start when we try to add on more helicopters later.

To find the best strategy with more than 5 bases, we first consider the approach of spreading out all the helicopters we have evenly and allocate the extra ones to the bases that are closer to where large numbers of calls tend to occur. The percentage of calls dispatched and the response time using this method are roughly 86% and 25.8 minutes. However, extremely low utilization rates around 7% indicate that we are 'wasting' helicopters. This could mean two things: one is that helicopters are being placed too sparsely, or it could also mean that we simply do not need all 12 helicopters.

We will explore the first possibility first, by placing helicopters to busier hospitals instead of spreading them all out to reduce 'waste'. However, the utilization rates of helicopters actually did not improve very significantly, which means that we probably have too many helicopters out there. Moreover, the average response time also went up by approximately 2 minutes. As a result, we further improve the model by reducing the number of helicopters to 11 and 10. And we find that having 11 helicopters actually increases the utilization rate by 5-10% while keeping other performance measures roughly the same. And having 10 helicopters will increase the utilization rate by 10-15%; however, the average response time will also go up around 3 minutes compared to when we had 12 helicopters. Overall, we discovered the trend that when helicopters are more evenly spread out, it is less cost-efficient and a lower percentage of calls will have helicopters dispatched. Nevertheless, it also tends to have a lower response time. On the other hand, when

helicopters are more centralized, they will be more efficiently utilized yet the response time might have to compensate a bit for this improvement.

As we decrease the number of helicopters, another problem also arises—-how can we optimize the location of these helicopters both for better results and a more feasible simulation process. After all, when we have more helicopters than the number of bases, it is somewhat common sense that it probably will not be a good idea to place all helicopters at one or two bases. However, when we have 7 helicopters, neither pure intuition nor pure computation will solve the problem. Therefore, we decided to look at it a bit more creatively and think of the problem not as randomly placing 7 helicopters, but more like adding another 2 helicopters based on our previous findings. Earlier we discovered an ideal combination for 5 helicopters and we will slowly increment the number of helicopters by taking this ideal combination as a default, then place the extra helicopters based on the locations of calls and heuristics.

Following this logic, the suggestions we will give when having more than 5 helicopters and an unlimited amount of bases will be (95% confidence interval):

#Helicopters	Based	Calls Dispatched	Avg. Response Time(hours)	Responded	Utilization of Helicopter
6	Ithaca, Binghamton, Buffalo, Rochester, Syracuse, Watertown	0.631 (0.6286, 0.6341)	0.536 (0.5336, 0.5382)	0.570 (0.5673, 0.5721)	9.50 (9.48, 9.52)
7	Ithaca, Binghamton, Buffalo, Rochester, Syracuse, Watertown, Elmira	0.670 (0.6682, 0.6724)	0.506 (0.5037, 0.5079)	0.606 (0.6036, 0.6081)	8.640 (8.620, 8.660)
8	Ithaca, Binghamton, Buffalo, Rochester, Syracuse, Watertown, Elmira, Albany	0.781 (0.780, 0.782)	0.487 (0.486, 0.488)	0.705 (0.704, 0.706)	8.810 (8.800, 8.820)
9	Ithaca, Binghamton, Buffalo, Rochester(2), Syracuse(2), Watertown, Elmira	0.805 (0.804, 0.806)	0.474 (0.473, 0.475)	0.727 (0.726, 0.728)	8.080 (8.079, 8.081)
10	Ithaca, Binghamton(2), Buffalo(2), Rochester(2), Syracuse(2), Watertown, Elmira	0.817 (0.816 0.818)	0.454 (0.453, 0.455)	0.738 (0.737, 0.739)	7.370 (7.360, 7.380)

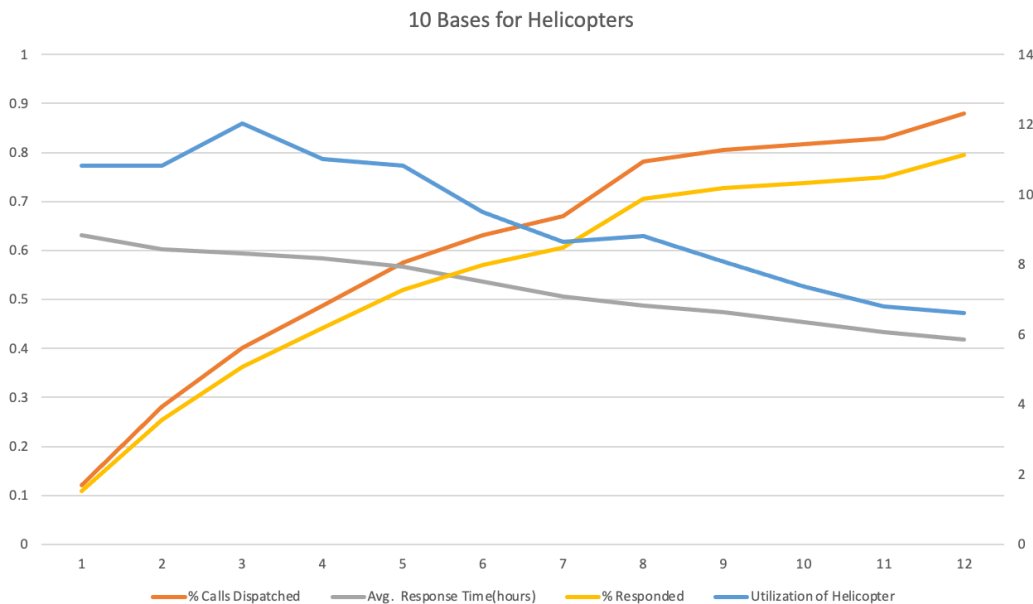| 11	Ithaca, Binghamton(2), Buffalo(2), Rochester(2), Syracuse(2), Watertown, Albany	0.829 (0.828, 0.830)	0.434 (0.433 0.435)	0.749 (0.748, 0.750)	6.810 (6.80, 6.82)
12	Ithaca, Binghamton, Buffalo(2), Rochester, Syracuse, Watertown, Sayre PA, Elmira, Albany(2), Utica	0.880 (0.879, 0.881)	0.418 (0.417, 0.419)	0.795 (0.794, 0.796)	6.620 (6.610, 6.630)

Figure 6: Results comparison when it has all bases available

One thing worth noting here is that the suggestions were not made based on one simple performance measure; instead, it is the result of many different factors. The general order we followed was to first look at the fastest response time, then examine the utilization rates, then look at the percentage dispatched, finally look at the number of bases being used. One result might have a slightly lower response time yet another result will have much higher utilization rates and percentage dispatched while only taking 2 more minutes on average for responses. When handling cases like this, we took into account real-life considerations and will choose the second option. However, an extensive list of results can be found in the appendix and be used to make different decisions.

Finally, when thinking about situations where we want to limit our spending on bases and only focus on five of them, it's important to strategically select these five locations. We decided to use the previous

finding and set that optimal combination (Ithaca, Buffalo, Rochester, Binghamton, Syracuse) as the default. Then add more helicopters to bases as needed. Using this logic, we acquired the following results (95% confidence interval).

| #helicopters	City Indices	Based	Calls Dispatched	Avg. Response Time(hours)	Responded	Utilization of Helicopter
6	6,7,9,0,5,9	Ithaca, Binghamton, Buffalo, Rochester, Syracuse(2)	0.633 (0.6319, 0.6341)	0.5421 (0.5420, 0.5422)	0.5715 (0.5704, 0.5726)	9.52 (9.50, 9.54)
7	6,7,9,0,5,6,9	Ithaca, Binghamton, Buffalo(2), Rochester, Syracuse(2)	0.6617 (0.6604, 0.6630)	0.5206 (0.5198, 0.5214)	0.5958 (0.5945, 0.5971)	8.51 (8.49, 8.53)
8	6,7,9,0,5,6,9, 5	Ithaca, Binghamton(2), Buffalo(2), Rochester, Syracuse(2)	0.6921 (0.6932, 0.6910)	0.4993 (0.4992, 0.4994)	0.6241 (0.6240, 0.6242)	7.8 (7.79, 7.81)
9	6,7,9,0,5,6,9, 5,7	Ithaca, Binghamton(2), Buffalo(2), Rochester(2), Syracuse(2)	0.7043 (0.7032, 0.7054)	0.4701 (0.4693, 0.4709)	0.6353 (0.6339, 0.6367)	7.06 (7.04, 7.08)
10	6,7,9,0,5,6,9, 5,7,0	Ithaca(2), Binghamton(2), Buffalo(2), Rochester(2), Syracuse(2)	0.7152 (0.7140, 0.7164)	0.4551 (0.4547, 0.4559)	0.6461 (0.6449, 0.6473)	6.46 (6.45, 6.47)
11	6,7,9,0,5,6,9, 5,7,0,5	Ithaca(2), Binghamton(3), Buffalo(2), Rochester(2), Syracuse(2)	0.7225 (0.7214, 0.7236)	0.4464 (0.4457, 0.4471)	0.6524 (0.6512, 0.6536)	5.93 (5.92, 5.94)
12	6,7,9,0,5,6,9, 5,7,0,5,6	Ithaca(2), Binghamton(3), Buffalo(3), Rochester(2), Syracuse(2)	0.7237 (0.7224, 0.7250)	0.4385 (0.4378, 0.4392)	0.6536 (0.6523, 0.6549)	5.45 (5.44, 5.46)
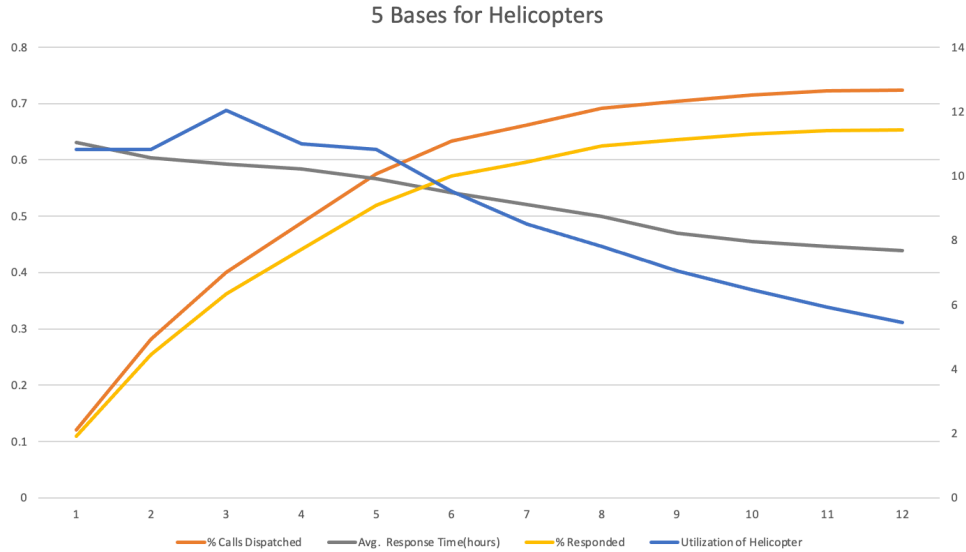
Figure 7: Results comparison when it has 5 bases available

Overall, we can conclude that considering response time, cost efficiency, and percentage dispatched, having 9-10 helicopters appears to be the most ideal. And we will place more helicopters at bases that are closer to more calls, instead of simply spreading all the helicopters evenly. This middle spot combines the advantages of having abundant helicopters—- fast response time, and more centralized placement—-much more cost-efficient.

Sensitivity Analysis

In building our model, we estimated certain parameters, such as call arrival rate, from real-world data. Due to the inherent randomness in these data, and the possibility of changes in these parameters in future applications, we need to do a sensitivity analysis to assess our model's robustness. Namely, we apply a small perturbation (10-20%) to certain parameters and observe the quantitative changes that this has induced in our model. Specifically, we are interested in the effects on the average response time and the utilization rate, since they describe the efficiency and cost-effectiveness of the model. We consider both scenarios where only up to 5 bases are allowed and where unlimited (maximum 10) bases are allowed. For each scenario, we test the model's robustness in three typical cases: when the number of helicopter = 1, 5, 12, which are the edge cases that will test the model's robustness the most.

We analyzed our model and determined that three parameters are the most important: the arrival rate of calls (lambda_max), the probability that a call is not safe (p_not_safe), and the rate of a call getting canceled on route (cancel_rate). The call arrival rate is significant because it may vary depending on seasons, months, etc., as we want our model to be able to handle scenarios when calls arrive more frequently than usual. The probability of unsafe calls is also prone to change: during snowstorms, thunderstorms, or other types of

inclement weather over certain regions, it may be more probable for a call to be deemed unsafe, so our model should be able to account for this. As for the cancellation rate, the cancel delay of a helicopter en route is assumed to be exponential; however, in our data analysis, the shape of the delay distribution seems ambiguous, which possibly implies uncertainty in the rate parameter. For other parameters of our model, we deem them sufficiently accurate so as to not have a major impact on the model's robustness. For example, the two gamma distributions modeling scene time and hospital time are shown to have excellent well-fitness in the Q-Q plots, so they can be safely assumed as accurate.

The detailed results of our sensitivity analysis can be found in Appendix 2. They demonstrate that, in general, our model is robust to perturbations. In all simulations of a variety of noise values applied to the parameters, the average response time and utilization are not significantly affected. The changes in their values are proportional to the changes introduced to the inputs. For example, a 10% increase in a parameter value may lead to about 5% difference in the performance metric. This shows that our model is well-equipped to handle unforeseen circumstances such as sudden spikes in the call arrival rate or severe weather causing a higher probability of unsafe calls.

Conclusion

From the above simulation results, several conclusions can be drawn. First, although intuitively it might seem like having as many helicopters as possible and spreading them evenly across the state will be the most efficient, it is actually not quite so. When we are not limited to a certain number of bases, having around 9-10 helicopters and placing them across 6-7 bases seems to be the option that balances all factors the best. The average response time is reasonably fast while ensuring the cost efficiency as much as possible. In general, when there are more helicopters than the number of bases, the performance measures are more ideal if we have the resources to allocate them freely and not be limited to five bases. And if we were to control the costs by limiting the number of bases to be 5, having 6-7 helicopters will maximum the cost efficiency which aligns with our goal the best. And it will also produce reasonable response time and percentage dispatched. To summarize, the results for optimizing placements for 1-12 helicopters with an unlimited number of bases and only 5 bases are described below.

#helicopters	Bases (unlimited)	Bases (n = 5)

1	Albany	Albany
2	Binghamton, Albany	Binghamton, Albany
3	Watertown, Binghamton, Albany	Watertown, Binghamton, Albany
4	Watertown, Binghamton(2), Albany	Watertown, Binghamton(2), Albany
5	Ithaca, Binghamton, Buffalo, Rochester, Syracuse	Ithaca, Binghamton, Buffalo, Rochester, Syracuse
6	Ithaca, Binghamton, Buffalo, Rochester, Syracuse(2)	Ithaca, Binghamton, Buffalo, Rochester, Syracuse(2)
7	Ithaca, Binghamton, Buffalo(2), Rochester, Syracuse(2)	Ithaca, Binghamton, Buffalo(2), Rochester, Syracuse(2)
8	Ithaca, Binghamton(2), Buffalo(2), Rochester, Syracuse(2)	Ithaca, Binghamton(2), Buffalo(2), Rochester, Syracuse(2)
9	Ithaca, Binghamton(2), Buffalo(2), Rochester(2), Syracuse(2)	Ithaca, Binghamton(2), Buffalo(2), Rochester(2), Syracuse(2)
10	Ithaca(2), Binghamton(2), Buffalo(2), Rochester(2), Syracuse(2)	Ithaca(2), Binghamton(2), Buffalo(2), Rochester(2), Syracuse(2)
11	Ithaca(2), Binghamton(3), Buffalo(2), Rochester(2), Syracuse(2)	Ithaca(2), Binghamton(3), Buffalo(2), Rochester(2), Syracuse(2)
12	Ithaca(2), Binghamton(3), Buffalo(3), Rochester(2), Syracuse(2)	Ithaca(2), Binghamton(3), Buffalo(3), Rochester(2), Syracuse(2)

Appendix 1

In this appendix, we give the details of the discrete event time simulation model we used to investigate the project's problem statement. We first give sufficient details to recover our implementation for a single replication of an experiment (with the experiment's controls being the number and location of the helicopters), and then give a detailed framework of the various experiments performed by using this developed simulation model for a single replication.

Running a single replication

- Initialization

The only inputs needed to run a single replication are the number of days for which the replication should run, the number of helicopters to consider, and the locations of said helicopters. All other states of the entities are randomly generated aside from the initial states of the helicopters which are all 'available' or, equivalently, ready to answer a call. We also instantiate performance metrics that will keep track of the times and values of various events produced throughout the replication. In particular, we keep track of the time at which unanswered calls occur and the response times to answered calls.

We first generate all the calls which will be used in the simulation for, say, a 3 month period, and populate the event list with these calls after which we proceed with the simulation logic as planned. We ask the reader to contrast this with generating calls on the fly where a call is generated only when the previous call is executed in the simulation logic. We opted for the former method for two main reasons:

(i) despite the computational burden of having to store all of the calls in the initial event list, the overall arrival rate of the call (computed as the time-averaged integral of the estimated arrival rate of a nonhomogeneous Poisson process over a 24-hour period) does not yield a substantial amount of calls to warrant generating the calls as the simulation deems them necessary.

(ii) A well-known fact from stochastic processes is that nonhomogeneous Poisson processes do NOT have exponential interarrival times, and, instead, require either an inversion method that has to compute an integral of the arrival process to obtain the next call time or another method often termed 'thinning' in which calls are generated over a predetermined time horizon and which are pruned according to a suitable probability function depending on the arrival process rate.

For these aforementioned reasons, we chose the latter method of 'thinning' to initialize the event list with all the calls for a particular replication.

- Main Iteration

We follow the general framework of a discrete event simulation by advancing the simulation clock to the time of the next event in the list and process said event by creating new events & inserting them into the event list and/or updating states of any relevant entities. There is undoubtedly a multitude of events to juggle around, and so, to these ends, we rely on various helper functions in our implementation which the interested

reader may or may not use him/herself. In particular, we constructed a total of six possible events along with their corresponding helper functions which we now describe:

(i) If the event is a 'call', simply schedule the time (according to an appropriate probability distribution function) at which a decision is made at HD.

(ii) If the event is a 'decision is made at HD', possibly deem the call unsafe or unavailable or canceled. If not, then the call can be answered with an appropriately chosen helicopter. Set the state of this helicopter to unavailable. A helper function finding the best available helicopter (where 'best' pertains to closest) may be of great use here.

(iii) If the event is a 'helicopter arrived at scene of call', simply schedule the time at which a decision is made at the scene.

(iv) If the event is a 'decision is made at the scene', find the 'best' hospital (in the same sense as above) while taking into account the possibility of a patient requiring special treatment.

(v) If the event is a 'helicopter arrived at hospital', simply schedule the time at which the helicopter returns to base.

(vi) If the event is a 'helicopter arrived at base', set the state of the helicopter to 'available'

It is worth mentioning that the most challenging events to the model were undoubtedly (ii) when a decision was made at HD and (iv) a decision was made at the scene of an answered call. The main difficulty lies in having to decide the best available option, whether it be a helicopter or hospital while retaining any inherent variability imposed by the model assumptions. We resolved these complexities by constructing a careful sequence of if/else statements that offered us both the flexibility to handle any call's particular realization of randomness and the capability of efficiently executing any relevant computations. For instance, in handling the possibility of a patient requiring special treatment (and thus requiring the nearest major trauma center as opposed to general hospital), we first considered whether or not the closest hospital was a major trauma center in which case our search would end. This is more than not the desired result since calls tend to be heavily populated around the major centers in which major trauma centers are located by default. Otherwise, we consider the small possibility the patient requires special treatment and assign them to an

appropriate hospital. It is through this clever use of simulation logic that we can handle similar cases of safety, availability, cancellation, etc. of calls in a clean and efficient manner.

- Termination / Collection of Relevant Performance Metrics

The replication ends when the simulation clock exceeds the specified time in the input parameters of the function. It is worthwhile to return back not only scalars indicating the average of various performance metrics embedded within a call, but also the times associated with these calls. This allows for great flexibility by allowing straightforward implementations of replication/deletion and batch means methods common in simulation practice. In particular, one may impose a warm-up period of arbitrary length by deleting or, rather, not considering the calls for which their recorded time is less than the warm-up period; in a similar vein, one may segment the time horizon into separate bins, as one does in batch means, and then conduct relevant statistical tests with calls received within each bin. Thus, we recommend that someone who is interested in implementing our outlined model to keep track of not only a single value of the performance metric associated with certain calls of interest, but rather keep a running list/dictionary of the times to allow for more sophisticated simulation analysis techniques such as replication/deletion and batch means.

Appendix 2

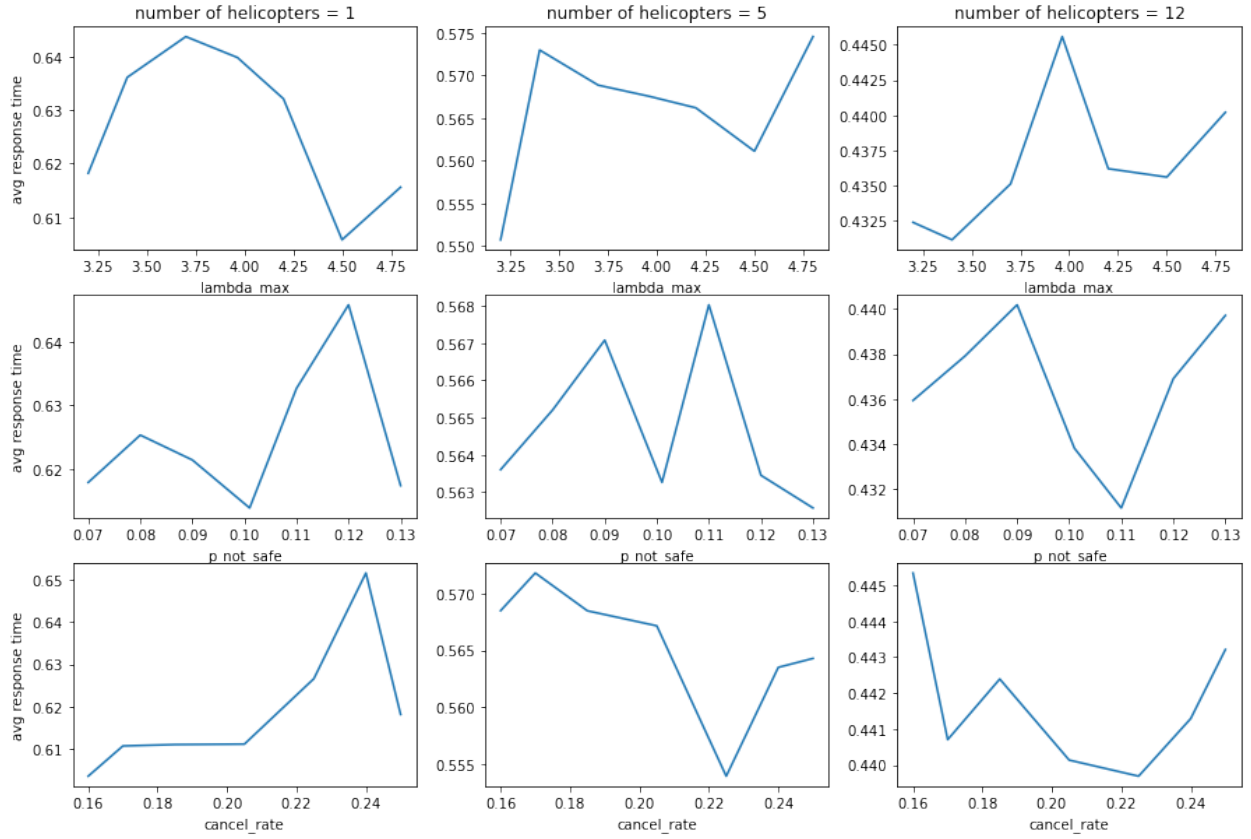In this appendix, we list the plots generated during the sensitivity analysis.

Figure 8: Change in average response time with perturbations in different model parameters. This is the scenario when we only use at most 5 bases. Subplots in each row correspond to perturbations in call arrival rate, probability of unsafe calls, and cancellation rate. Subplots in each column correspond to different numbers of helicopters. The figures below follow the same format

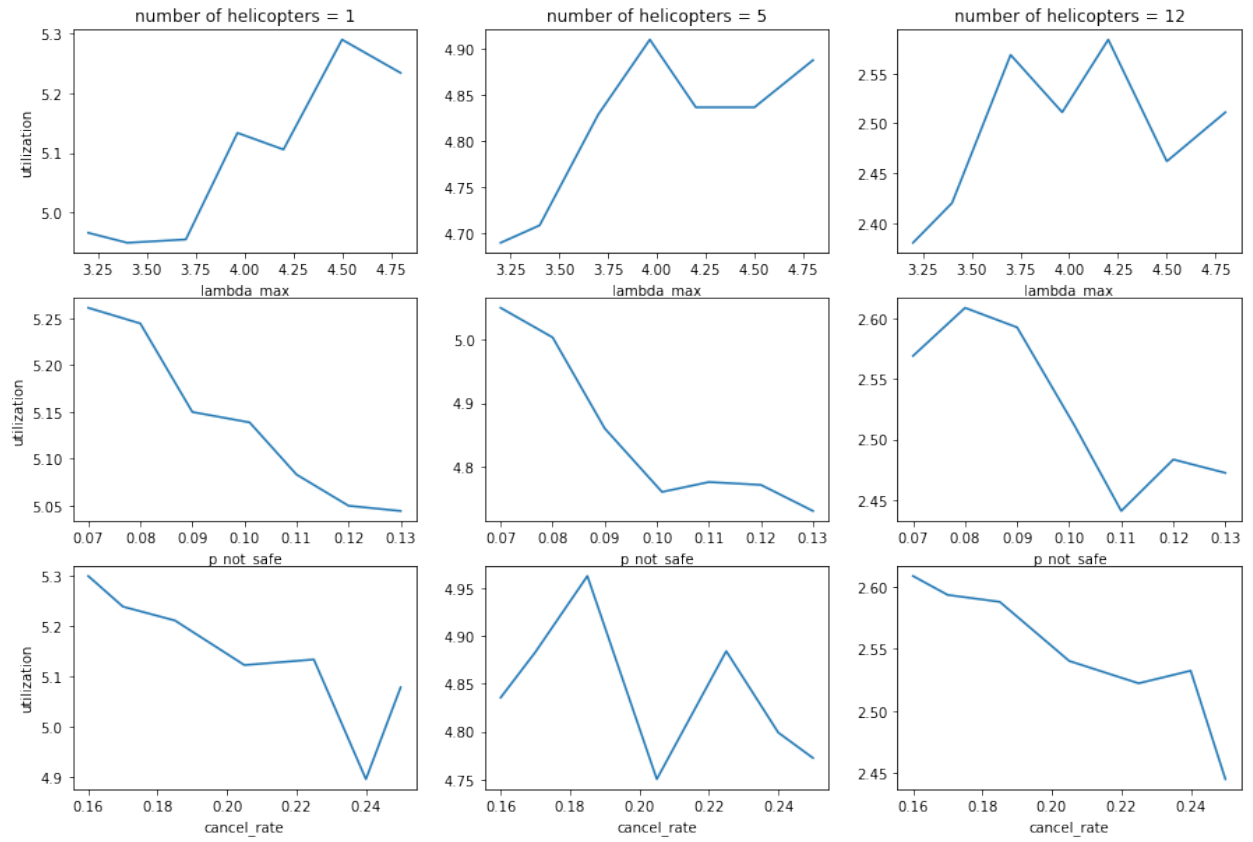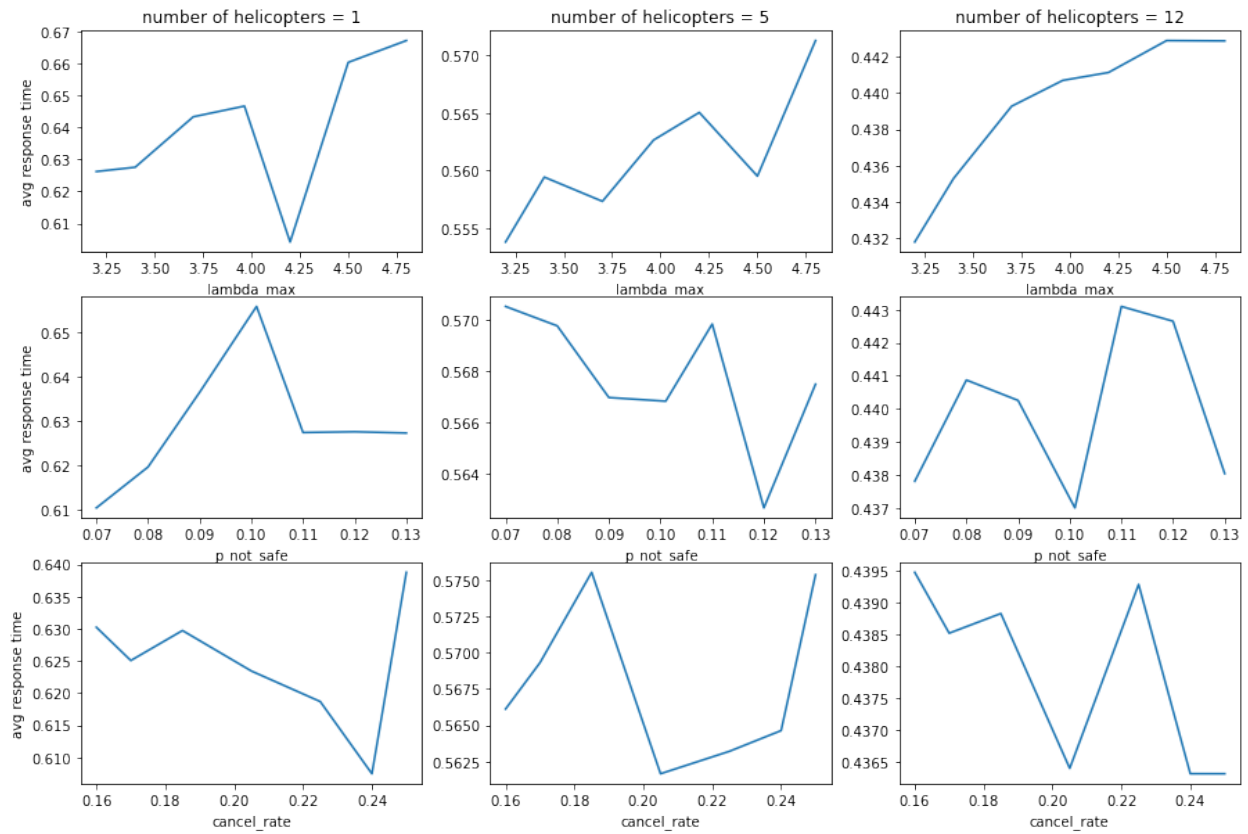Figure 9: Change in utilization with perturbations in different model parameters. This is the scenario when we only use at most 5 bases.

Figure 10: Change in average response time with perturbations in different model parameters. This is the scenario when we can use at most 10 bases
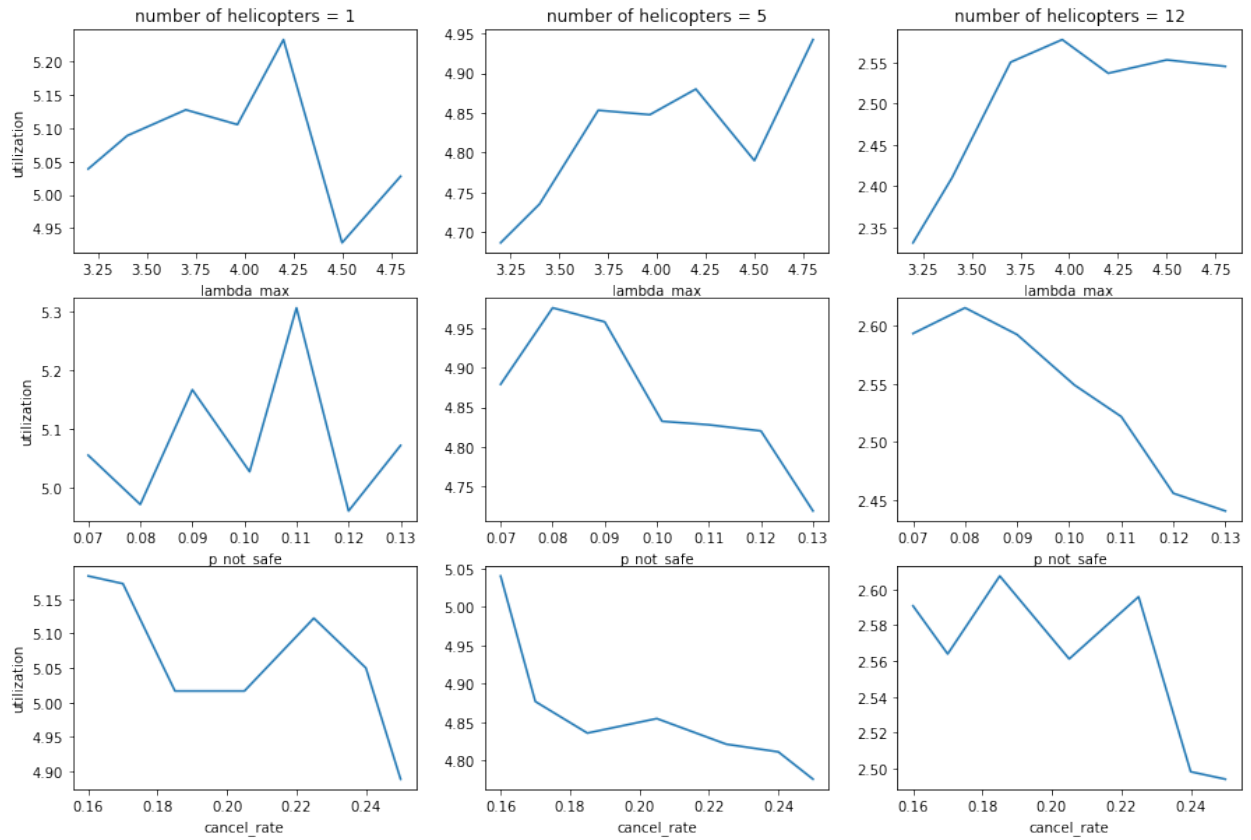
Fig 11: Change in utilization with perturbations in different model parameters. This is the scenario when we can use at most 10 bases