Optimizing Emergency Response Crew Allocation during Earthquakes to Improve Restoration Time

by

Mohamed Kurdi
B. Computer Engineering, McGill University 2013

Submitted to the MIT Sloan School of Management and the Institute for Data, Systems, and Society in partial fulfillment of the requirements for the degrees of

Master of Business Administration
and
Master of Science in Engineering Systems

In conjunction with the Leaders for Global Operations Program at the
Massachusetts Institute of Technology
June 2017

Signature of Author .... Signature redacted
MIT Sloan School of Management
Institute for Data, Systems and Society
May, 12, 2016

Certified by .... Signature redacted
Dr. Georgia Perakis
Thesis Supervisor, William F. Pounds Professor of Operations Research

Certified by .... Signature redacted
Dr. Saurabh Amin
Thesis Supervisor, Robert N. Noyce Career Development Assistant Professor

Accepted by.................. Signature redacted ....
John N. Tsitsiklis
Clarence J. Lebel Professor, Electrical Engineering
IDSS Graduate Officer

Accepted by.. Signature redacted ....
Maura Herson
Director of MBA Program MIT Sloan School of Management

THIS PAGE INTENTIONALLY LEFT BLANK

Optimizing Emergency Response Crew Allocation during Earthquakes to Improve Restoration Time
by
Mohamed Kurdi

Submitted to the MIT Sloan School of Management and the Institute for Data, Systems, and Society in partial fulfillment of the requirements for the degrees of
Master of Business Administration
and
Master of Science in Engineering Systems

Abstract

Northern California and specially the San Francisco Bay Area where PG&E operates is very susceptible to earthquakes. United States Geological Survey (USGS) estimates a 63 percent chance that a magnitude-6.7 or larger earthquake will hit the Bay Area by the year 2036. The chances for a 7.0-magnitude or above are about 50 percent.

In this thesis, we first present the methodology PG&E uses to generate predicted damages. Then, we will discuss what data will be available to us and outline how this data is transformed into predicted damages for pipes.

Then, the thesis go over the method we used to generate the predicted customer service calls per area. It will first present how PG&E currently estimates the number. Then, it will present a model that can provide better accuracy for estimating the numbers.

Next, we present a resource allocation model to optimize repair crew allocation between divisions. We will present how the resource allocation problem can be formulated as a load-balancing problem. We present different formulations and discuss the run time and benefits/drawbacks of each model. We formulate a two-stage optimization model and a one-stage optimization model. We ran both models on different scenarios and we compared the results. We also highlight some key insights we got from combining the travel and allocation problem in a single stage optimization problem.

We also go over the sources of uncertainty we have in our data. There are three sources of uncertainty in the model. In this thesis, we will model one of the sources of uncertainties and outline how the other two can be incorporated into the model in the future.

Finally, we generated ideal outputs for some of the likely USGIS scenarios that PG&E includes in their emergency response plan.

The results from this model would be a critical input to PG&E's emergency response team during an earthquake event. The better we are at predicting damage and allocating resources, the better we will be at minimizing earthquake impact on communities.

Thesis Supervisor: Georgia Perakis
William F. Pounds Professor of Operations Research
MIT Sloan School of Management

Thesis Supervisor: Saurabh Amin
Robert N. Noyce Career Development Assistant Professor
Department of Civil and Environmental Engineering

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgment

I would like to thank my advisers Prof. Georgia Perakis and Prof Saurabh Amin for their guidance and support on this project.

The team I worked with at PG&E was supportive and helped me tremendously during my time at PG&E and after I left. Thank you Bryan Hennessy, George Gaebler and Nick Raad for guiding me through the operations of PG&E and welcoming me into the team.

I would also like to thank my fellow LGO and Sloan classmates for the unforgettable experience at MIT and Sloan.

Finally, I would like to thank my friends, family, my wife and my son, for their encouragement, love and support during my time at MIT

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

1 Introduction

1.1 Overview and Project Motivation

PG&E's gas operations supports over 4.3 million natural gas customer accounts, and the gas infrastructure consists of 48,000 miles of gas pipeline. Northern California and specially the San Francisco Bay Area where PG&E operates is very susceptible to earthquakes. United States Geological Survey (USGS) estimates a 63 percent chance that a magnitude-6.7 or larger earthquake will hit the Bay Area by the year 2036. The chances for a 7.0-magnitude or above are about 50 percent.

In recent years, PG&E has made substantial investments in modeling technology to enable the company to develop resiliency plans, estimate resource needs, and respond more quickly to keep the public safe. In cases like the Napa County earthquake, which hit on Aug. 24 2014, the models helped deploy PG&E's field teams to conduct inspections and make emergency repairs.

PG&E's current earthquake damage model can intake seismic event data, and produce the number of expected failures in each approximately ¼ square mile size land segment in the region. The traditional way of organizing response efforts has been to send available resources based on most probable damage predicted from these models. The goal of this project is to assess whether it is possible to improve the damage perdition and to improve the resource allocation after an emergency.

The results from this model would be a critical input to PG&E's emergency response team during an earthquake event. The better we are at predicting damage and allocating resources, the better we will be at minimizing earthquake impact on communities.

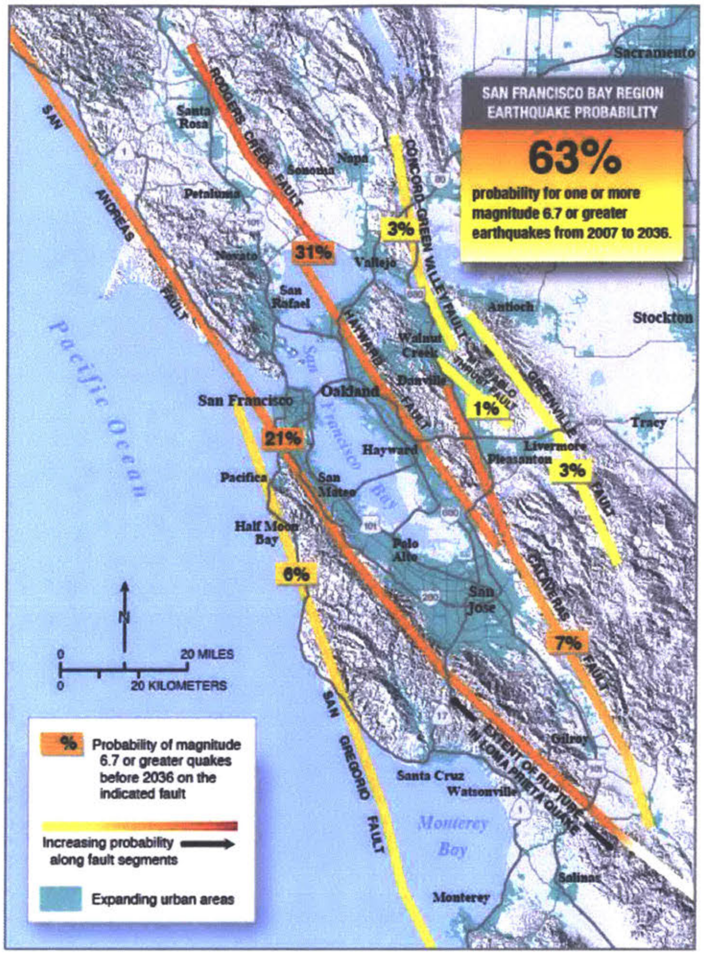


Figure 1-1: The map shows the major faults and the odds that they will cause a major earthquake

1.2 Current Earthquake Emergency Response Process

The process PG&E currently uses for determining the resources required is semi-manual and depends on operators' and planners' accumulated knowledge. After an earthquake hits, initial damage estimates for transmission pipelines and distribution plats will be provided by DASH (Dynamic Automated Seismic Hazard). Personnel will use the DASH output to estimate shaking, damage at PG&E facilities and assets. DASH output facilitates rapid identification of potential problem areas for gas assets prior to the receipt of damage reports from the field. Using the DASH report estimates, the number of damages estimated and the average repair times, the planning team calculates the number of crews required to restore the system and decide whether they need to request extra crews from other utilities through mutual aid agreements.

The restoration efforts are managed from the Gas Emergency Center (GEC). The team at the GEC keeps track of discovered leaks, leak reports and the location of the repair crews. Using the output of the DASH report, the team creates a plan for the leak survey crews to survey the entire affected area based on likelihood of damage estimated by DASH. At the same time, they manually assign workers to active leaks that have been identified already. The process is summarized in the figures below:

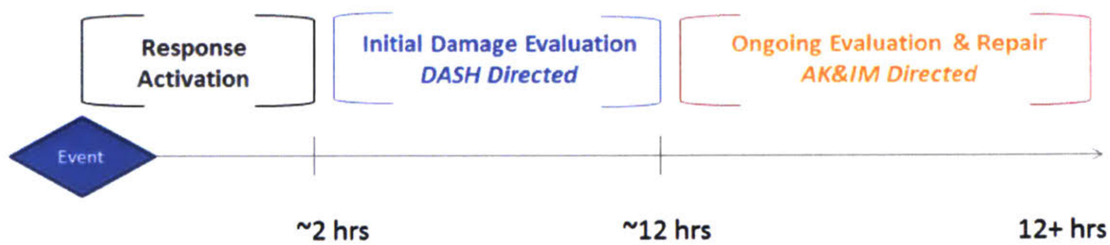


Figure 1-2 Emergency Response Timeline

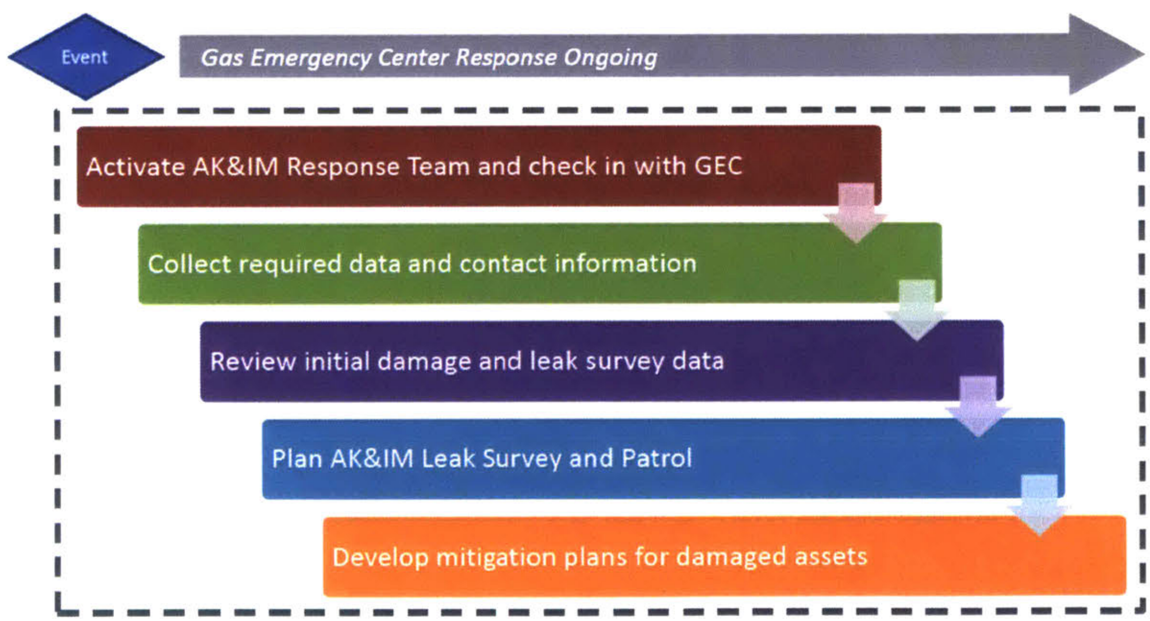


Figure 1-3 Emergency Response Flow Diagram

There is not official criteria to decide when the systems has been restored. For the purpose of this project and based on the interviews we have conducted, the system is restored after the following tasks are completed:

1- All areas marked by the DASH report and the Integrity Management team are surveyed for leaks. This task is performed by Leak Survey

2- A field service member ensured that all customer reported incidents (dispatch calls) are made safe. This task is performed by Gas Service Reps. (GSRs)

3- Relight jobs are completed. This task is performed by GSRs as well

4- All grade 1 leaks are repaired. This task is performed by the Maintenance Crews

Tasks 1 and 4 have their own dedicated crews to do the job. Tasks 2 and 3 are completed by the same crew type. Since tasks are performed by different crews, the planning for each crew type is independent of the other crew types.

The model is expected to produce the following:

a. Expected restoration finish time for each task (tasks 2 and 3 are pooled together)

b. Crew allocation by division

c. Where the crews should come from. For example, San Francisco should get five GSRs from San Jose yard and San Jose should get two GSRs from Oakland.

1.3 Thesis Overview and Contribution

In this thesis, we will first present the methodology PG&E uses to generate predicted damages. Then, we will discuss what data will be available to us and outline how this data is transformed into predicted damages for pipes.

Then the thesis will go over the method we used to generate the predicted customer service calls per area. It will first present how PG&E currently estimates the number. Then, it will present a model that can provide better accuracy for estimating the numbers. We also highlight how to improve the models if more earthquake data is available in the future. Our major contribution in this section is studying which factors predict odor calls and how to estimate the number of odor calls PG&E expects to receive after of an earthquake. When more data is available, PG&E can use the same methods to estimate the number of calls rather than the static method they currently use.

In Chapter 4, we will present how the resource allocation problem can be formulated as a load-balancing problem. We will present different formulations and discuss the run time and benefits/drawbacks of each model. More specifically, we create a two-stage optimization model and a one-stage optimization model. We ran both models on different scenarios and we compared the results. We also highlight some key insights we got from combining the travel and allocation problem in a single stage optimization problem.

We will also go over the sources of uncertainty we have in our data. There are three sources of uncertainty in the model. In this thesis, we will model one of the sources of uncertainties and outline how the other two can be incorporated into the model in the future.

Finally, we will generate ideal outputs for some of the likely USGIS scenarios that PG&E includes in their emergency response plan.

We have three contributions in the resource allocation section. First, we formulated the problem to with earthquakes emergencies in mind. We also ran the model on multiple earthquake scenarios and presented the results. Second, we present a way to combine the two stages of optimization (allocation and routing) into a single-stage formulation that takes into consideration the travel times when allocating jobs to crews. Finally, compared to the model presented by Whipple [10], we assign jobs to crews rather than divisions. Given how PG&E assign jobs, this formulation is more applicable.

Better allocation of resources will allow PG&E to reduce its restoration time. This will allow PG&E to mitigate the risk of earthquake impact. Additionally, reducing restoration time is critical to improve PG&E's public image and show the regulators that PG&E is well prepared for natural disasters such as earthquakes.

2 Literature Review

2.1 Damage Prediction in Pipelines Infrastructure

The ALA (American Lifelines Alliance) [1] presented procedures that can be used to evaluate the probability of earthquake damage to water transmission systems. They created a model that estimates that repair rate per unit length of pipes as a function of ground shaking. They identified different parameters that determines the relation between ground shaking and the repair rate such as pipe age, pipe material and soil corrosiveness.

Pineda and Najafi [2] examined the seismic damage estimation models for buried pipelines developed over the last three decades. PGD (Peak Ground Displacement) is what physically causes the damage. However, it is not easy to estimate given the current widely deployed sensor technologies. Moreover, it is well documented that PGV (Peak Ground Velocity) is a convenient parameter that is very easy to estimate and well correlated with pipeline damage. We will examine the relationship between PGV and odor calls as a proxy for leaks

Most the damage models were originally developed for segmented pipelines. Most of the damages happen at the joins. HAZUS-MH developed by FEMA [3] suggest that to estimate the damages for continuous pipelines, we use the same fragility relations of segmented pipelines multiplied by 0.3. However, their approach is likely overestimating the damages [2].

2.2 Optimization Literature Review

In emergency response, optimization concepts are not a novel approach and have been used in the industry for a long time. Many papers addressed crew scheduling and crew placement. Yao et al. [4] develops a model for pre-staging crews and then dispatching the once faults are discovered. A penalty is applied if crews can not reach the repair site within a the target repair period and the model attempts to minimize the penalty. Guha et al. [5] develops two algorithms for crew assignment problems. The first version allows customers to have different priority and tries to minimize the cost incurred by high priority customers. The second tries to minimize the restoration time for the entire system.

When dealing with uncertainty, many papers use stochastic or robust approaches to deal with the uncertainty. Herrolean and Leuus [6] provide a good review of the various techniques available for scheduling under uncertainty. Balwani [7] showed how a stochastic optimization can reduced overtime repairs at gas utility by looking at multiple classes of resources. Bertismas-Sim [8] developed an algorithm using uncertainty sets that takes into account the uncertainty in repair times. The paper provides a way for tradeoffs between robustness and performance. They also showed that the formulation is also valid for MIP [9]

Fiedrich and Gehbauer [11] developed an approach that optimized resource assignments by modeling the dynamic aspects of responding to earthquake emergencies during the early search-and-rescue operations. This search-and-rescue is analogous to PG&E's locate-and-cap part of the earthquake.

In this thesis, we modeled the problem as a mixed integer program following a similar model developed by Whipple [10]. A major limitation of the first model we present in this paper is that it assumes there are no uncertainties in the data set. We then took an approached developed by Bertsimas and Sim [10] to create a trade-off between robustness and optimality, which also extends to Mixed Integer Programs [10].

3 Damage Prediction

3.1 Overview of the Damage Prediction and Data PG&E Currently Uses

In terms of data capability, PG&E have been rapidly improving over the last few years. However, given the amount of data they have, there is more to be desired when it comes to creating insights from data.

PG&E uses DASH to predict the likelihood of leaks and the number of leaks in every area immediately after an earthquake, as we will describe later in this chapter.

Right now, PG&E does not have the capability to predict the expected number of customer calls. When allocating resources, the assumption is that the number of calls they get in an area is proportionate to the leaks in an area. This assumption is used when planning for resources just after an earthquake event hits.

PG&E uses a tool called TAMI (Tactical Analysis Mapping Integration) to plot the customer calls on a map. They can also see where resources are available on the same map and send resources to respond to customer calls in real time.

The screenshot below show's TAMI's view during the Napa earthquake. The green pins represent customer calls that have not been resolved yet. The contour plot shows the intensity of the earthquake and the red diamond in the most-inner area is the epicenter.

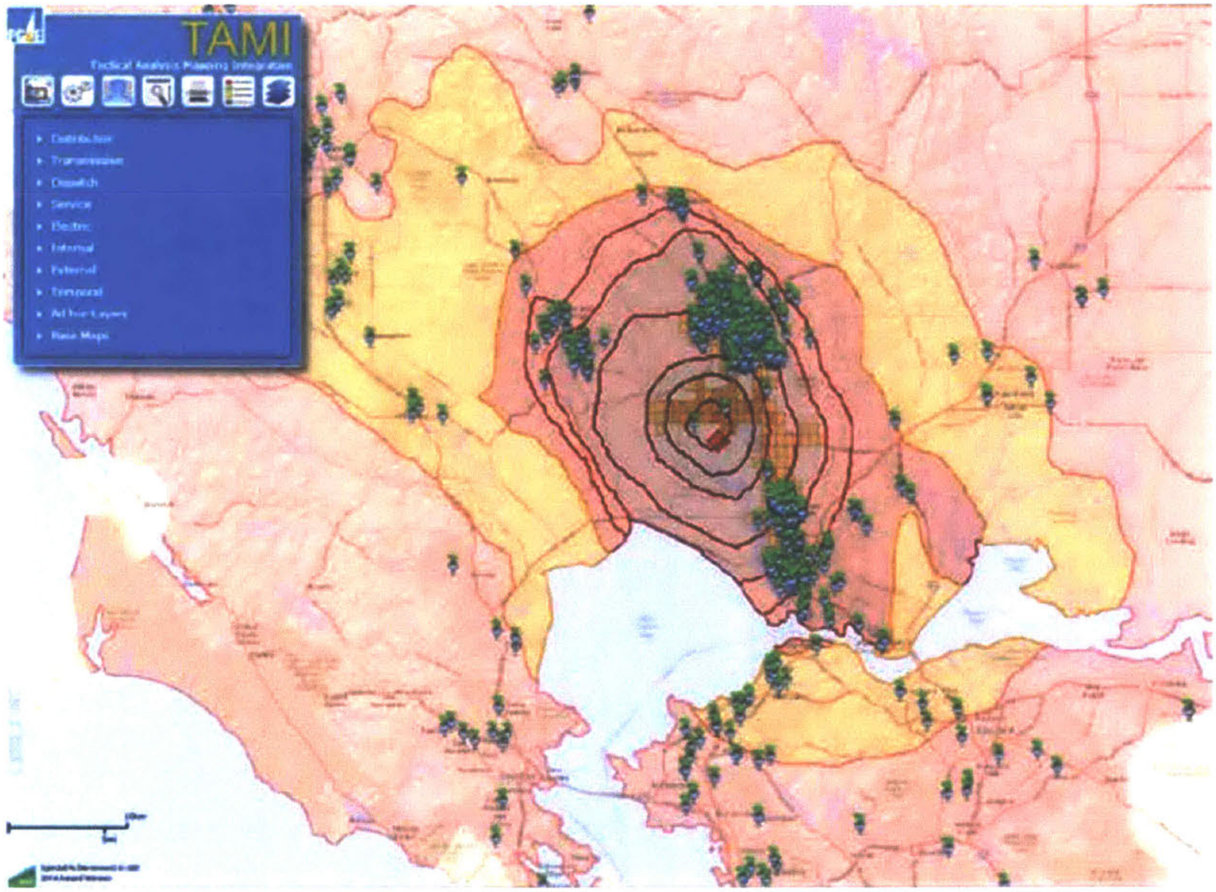


Figure 3-1: Odor calls overlaid on top of Napa earthquake shake map

Using TAMI, PG&E predicts where leaks might be. If many customers call and report gas odors in a limited geographical area, they create an event (called area odor call) and send leak survey crews to that area. Odor calls are believed to be an indicator of a leak. This approach relies heavily on the planner's knowledge and their knowledge of the assets located in areas with potential leaks.

PG&E still have to send a GSR to respond to those calls within a reasonable amount of time (30-60 min during normal operating procedures). During emergencies, the respond to less critical calls might be delayed, as resources are limited. A major limitation of PG&E's current approach to customer calls is that they respond to them after the fact. GSRs can be better utilized and positioned if we can predict customer calls in advance.

3.2 Data Available

The only major earthquake PG&E had some data on was the Napa 6.0 earthquake. There are two types of damages we are trying to predict. The first one is the leaks in the PG&E transmission and distribution network. The second is the customer calls that a GSR needs to attend.

The Napa earthquake only caused seven grade-1 leaks (the only type of leaks we are concerned with during emergencies) and most of them were caused by buildings moving off the base and breaking the riser coming off the ground. Thus, it was hard to build a prediction model to predict the leaks.

Another intuition we had was to use a similar process to what PG&E uses. We can look at customer odor calls as a proxy for leaks and establish with the few data points we had if there would be at least a correlation between the leaks and the density of customer calls. However, with those few data points, we could not find any statistically significant correlation. Thus, we had to find a different way to get an estimate that we can use in the optimization model during different expected earthquake scenarios.

Given the limited data we have about leaks, we used the DASH repair rate (the number of leaks per mile of pipes) to calculate the expected number of repairs in a given area. However, due to the lack of major earthquake in the last 25 years, we could not verify the estimate.

As for the customer calls, there were enough calls to build a model estimating the expected number of calls coming from an area affected by the earthquake. To construct our model, we utilized PG&E assets data, workers data, and leak repairs data. We also used the Napa 6.0 earthquake dispatch data to predict the number of dispatch calls we expect. Finally, we also used the USGS earthquake scenarios to build create scenarios which we can use to run and test our model.

3.3 DASH Overview and Leaks Estimates

The current gas pipeline repair estimates is based on a methodology developed for predicting water pipeline repair rates based on empirical data. It was developed by the American Lifeline Alliance [1]. While the water and gas pipeline share many features, some fundamental differences exist (e.g. gravity-fed vs pressurized systems); model parameters have been revised in order to reflect those differences when possible.

The three major components for the damage estimate methodology are inventory data, seismic hazard data and pipeline damage models.

Pipeline Data: The inventory was provided by PG&E. It includes information about 50,000 (7200 miles) pipeline segments and about 22,000 distribution plat sheets (23,000 miles of pipes).

Seismic Hazard Data: The hazard data include estimated patterns of ground shaking provided by USGS as well as estates of ground displacement resulting from liquefaction, landslides and fault ruptures.

Pipeline Damage Models: The function form of repair rates is represented in equations 1 and equation 2 for ground shaking and ground failure respectively

$$Repair\ Rate = K_1\ x\ 0.00187x\ PGV \tag{3-1}$$

$$Repair\ Rate = K_2\ x\ 1.06\ x\ PGD^{0.319} \tag{3-2}$$

Peak ground acceleration (PGA) is equal to the amplitude of the largest absolute acceleration recorded on an accelerogram at a site during a particular earthquake.

Peak ground velocity (PGV) is the greatest speed (rate of movement) reached by the ground. PGV merely expresses the peak of the first integration of the acceleration record

Permeant ground displacement (PGD) is ground movement during an earthquake. PGD is estimated by double integrating the PGA.

K_1 and K_2: constant scaling factors to reflect the differences when compared to baseline pipes

After an earthquake hits, PG&E uses the repair rate and aggregate the expected number of leaks over a division to come up with the leaks distribution. The number of dispatch calls is assumed to be ten times the expected number of leaks. In the next section, we provide an alternative mechanism that can be used to better predict the number and the location of dispatch calls.

3.4 Model Formation and Implementation for Estimating the Number of Dispatch Calls

As mentioned in the previous section, the number of dispatch calls is assumed to be ten times the number of expected leaks. We wanted investigate how to better estimate the number of dispatch calls. In this section, we go over multiple models to estimate the expected number of dispatch calls received due to an earthquake.

To start building the model, we first wanted to know which calls are caused by the earthquakes. We started by looking at calls in areas that were impacted by the earthquake. We created a baseline case of how many calls do we expect in a normal day at each hour of the day. We then then looked at the times where the number of calls was significantly higher than what we expect in a normal day to determine how many hours after the earthquake we want to consider when counting the dispatch calls received for a normal day.

By looking at how many calls we revived per hour, we can tell the calls are indeed caused by the earthquake. For example, during the early morning hours (3 – 7 am), we do not see many calls during a normal day. However, during Napa, which started around 3:20 am, we can clearly see the spike in the number of received calls. We see another spike at 8 am when people wake up and see that an earthquake happened last night (or potentially find a leak and call)

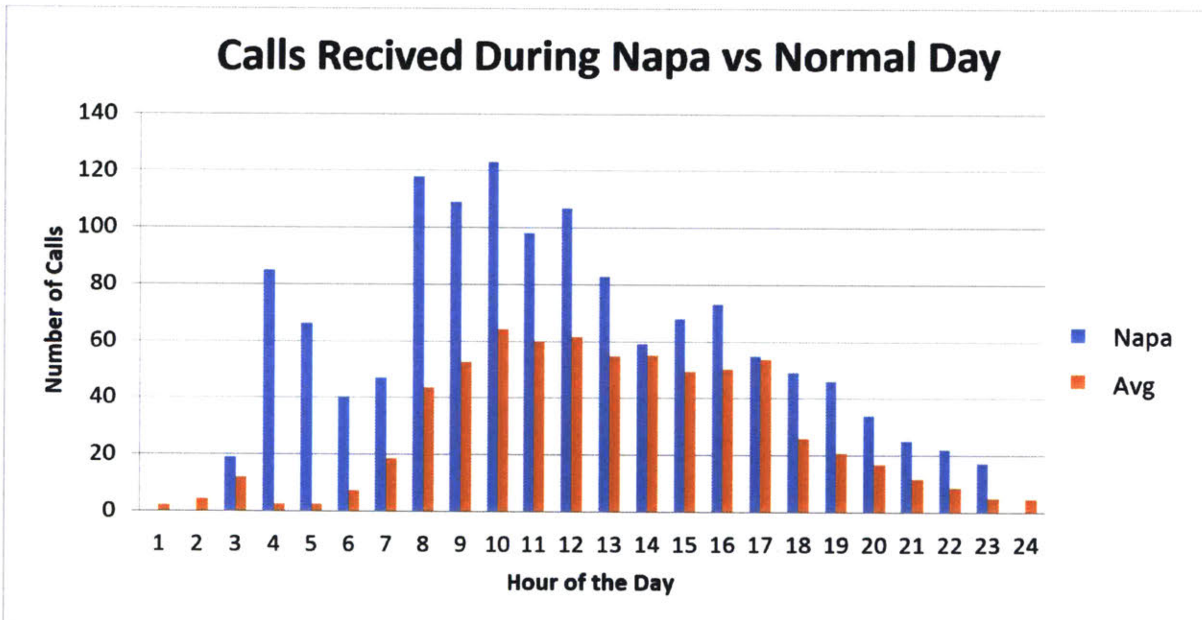


Figure 3-2: Customer calls received during Napa earthquake compared to a normal day

Using the NAPA earthquake data, we divided the affected area into a 0.25 km x 0.25 km squares. We then tried to correlate the dispatch calls density with the intensity of the earthquakes in those areas. We divided the data set into the training dataset and a testing dataset. We used the training dataset on multiple algorithms to create the model.

We first started by trying to create a simple regression model between the Peak Ground Velocity (PGV) which represents the earthquake intensity and the number of dispatch calls received in an area. The results of this model are show below:

Dependent Variable:

#Dispatch Calls: DC

Independent Variables:

Earthquake Intensity (represented as Peak Ground Velocity PGV): PGV

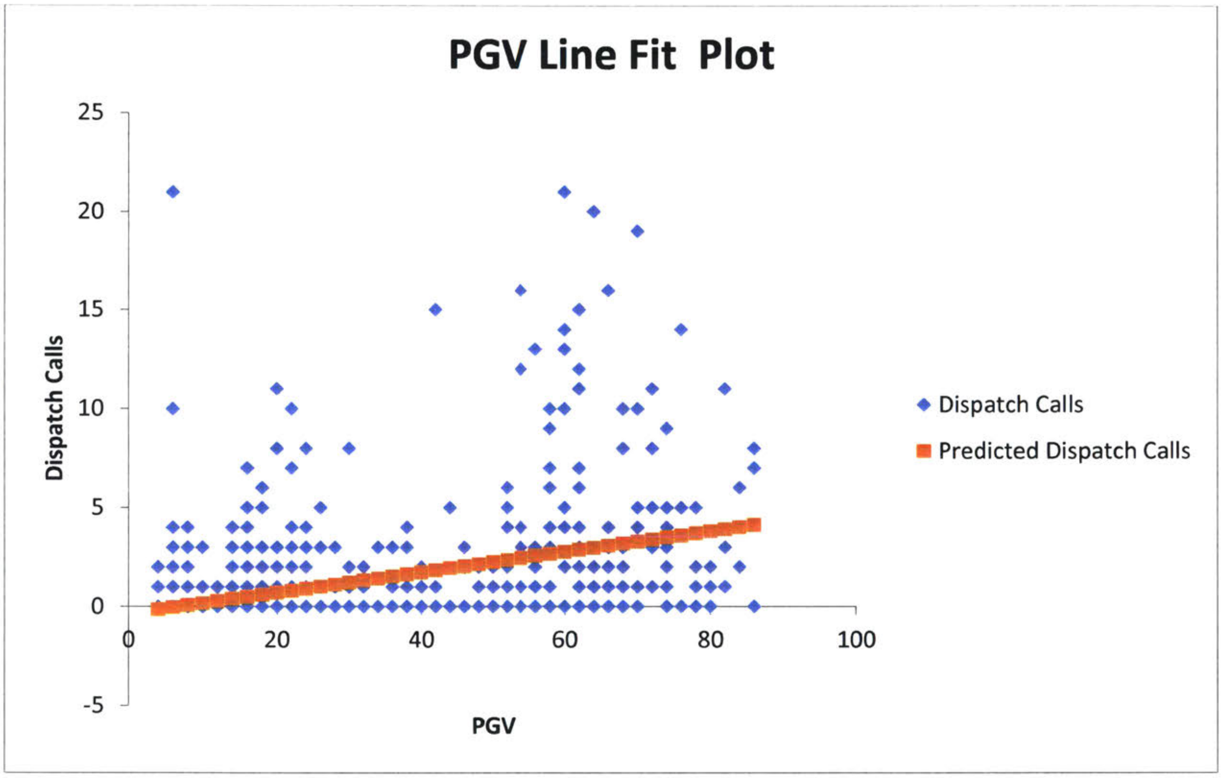


Figure 3-3: Correlation between earthquake intensity and the number of dispatch calls

Regression Statistics	
Multiple R	0.45
R Square	0.20
Adjusted R Square	0.20
Standard Error	1.91
Observations	1511.00

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1352.65	######	372.67	2.20E-74
Residual	1509	5477.15	3.63		
Total	1510	6829.81			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.321	0.072	-4.472	0.000	-0.461	-0.180	-0.461	-0.180
PGV	0.052	0.003	19.305	0.000	0.046	0.057	0.046	0.057

It was clear from the results that the earthquake intensity does correlate to the amount of calls we get. However, it was also clear that there were other missing factors. Given that customers initiate the dispatch calls, we decided to introduce the customer's density within that area as an independent variable. This model looks at the effects of both the customer density and the intensity of the earthquake. In the graph below, the color of the square represents the customers' density, the counter plot represents the intensity of the earthquake and the dots represents the location of the dispatch calls PG&E received up to 8 hours after the earthquake.

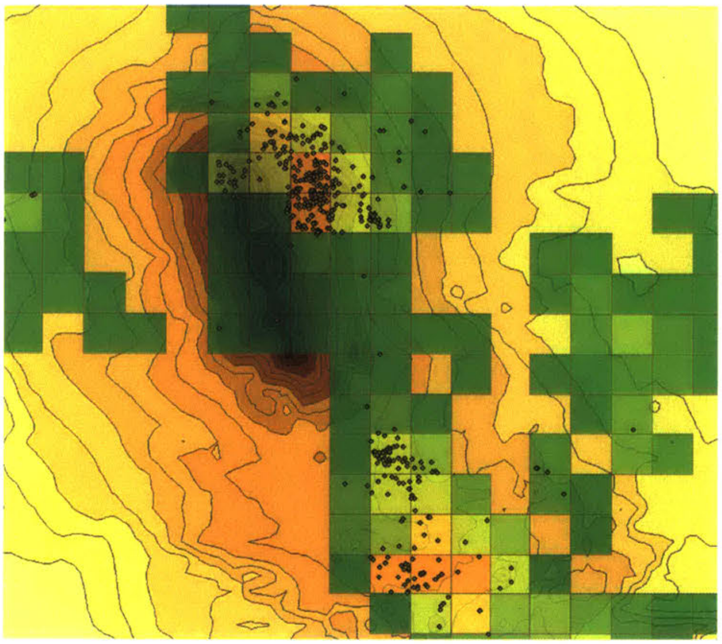


Figure 3-4: Customer's density, dispatch calls and earthquake intensity overlaid

The model is formulated as follows:

<u>Dependent Variable:</u>

#Dispatch Calls: D

<u>Independent Variables:</u>

Number of Customers: NC

Land Slide Susceptibility: LS (between 0 – 5)

Peak Ground Velocity: PGV

Soil Corrosiveness: CR

The regression output is shown below:

```
Linear regression model:
    DC ~ 1 + LS + CR*PGV + NC*PGV

Estimated Coefficients:
```

	Estimate	SE	tStat	pValue
(Intercept)	-0.80312	0.30648	-2.6205	0.0088682
LS	0.069426	0.032602	2.1295	0.033378
CR	0.23182	0.098314	2.358	0.018503
NC	-0.0025837	0.00056008	-4.6131	4.3043e-06
PGV	0.048099	0.023644	2.0343	0.042095
CR:PGV	-0.016831	0.0076699	-2.1945	0.028354
NC:PGV	0.00045192	1.9693e-05	22.949	3.5945e-100

```
Number of observations: 1511, Error degrees of freedom: 1504
Root Mean Squared Error: 1.5
R-squared: 0.502,  Adjusted R-Squared 0.5
F-statistic vs. constant model: 253, p-value = 1.34e-223
```

While the model above had relatively good results, some the parameter estimates don't follow with what one might expect. For example, the number of customer is negatively correlated to the expected number of calls even though we would expect the correlation to be positive. To further investigate that, we looked at the effect of the number of customers at different PGV values. The figure below shows the effect of the number of customers is insignificant when PGV is low. This makes intuitive sense since we don't expect more calls in areas that were not affected by the earthquake.

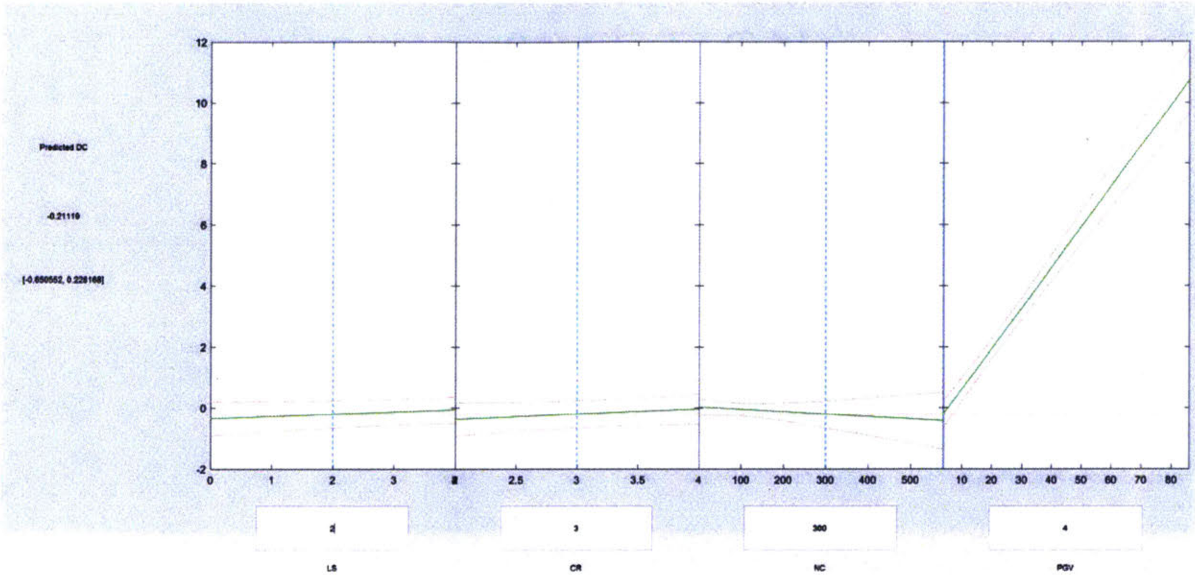


Figure 3-5: The effect of the number of customer's when PGV is low

3.5 Results

The table below shows summary results on the training and testing data sets from the different models we ran. See we describe the models in more details in appendix A

Method	R-Square Train	R-Square Test
Regression (No interactions)	0.33	0.22
Regression (Pair-wise interactions)	0.50	0.49
Neural Network	0.61	0.50

We can see that the pairwise interaction models performs much better than the model with no interactions. Looking at the data, the variable that explains most of the variance in the number of calls is the pair-wise interaction between the earthquake intensity and the customers' density. This intuitively makes sense. Even if an earthquake hits an area hard, there will not be many calls if the area has a very small number of customers.

3.6 Conclusion and Future Work

We showed that the amount of dispatch calls received from an area is correlated to the intensity of the earthquake in that area. Given the earthquake used to generate the data had a relatively very low impact, it is not clear if those calls would help in finding major leaks in a more severe event or if the number of calls would be exponentially higher in a major event.

In addition, we suspect that there is a psychological factor as well. We believe that people are more likely to report a leak if they feel an earthquake even if there was never a leak to begin with. This can be seen from the Napa data. Most of the reported leaks were not actual leaks.

It would be helpful to revisit the relationship in the future if more earthquake data is available. We believe that a relationship between the intensity of the earthquake and the amount of customer calls PG&E gets can be established. Furthermore, the dispatch calls can be used as a proxy to help locate major leaks early on during the emergency response effort. This would decrease the risk of major incidents significantly. Finally, in the future, this model can be used as an input to the resource allocation model during emergencies to improve GSRs allocation. This could give better results than the current static assumption of dispatch calls to expected leaks ratio.

4 Resource Allocation Model

4.1 Introduction to Emergency Restoration Tasks

During earthquakes, PG&E stations crews at sites called base camps. The crews operate from those base camps to serve the divisions they are assigned to. Those locations can be PG&E facilities or can be also municipal locations such as parks or schools where PG&E can temporary stage crews during the restoration period. PG&E's main goal is to improve the safety of the system by repairing leaks and responding to customer calls quickly. They are also interested in reducing service interruptions and improving their public image.

Currently PG&E uses a manual process to handle crew assignment. The emergency response team uses their best educated guesses to determine how many crews to assign to each division. Those decisions are likely to be sub-optimal specially since they usually lack experience to deal with earthquakes since most of them weren't working at PG&E during the last major earthquake. An understanding of what damages to expect and how to balance the workload between different divisions will aid the emergency response team in making restoration plans that ensures the system is restored in the shortest time possible.

As we mentioned in the process overview section, we consider the system to be restored after 3 types of tasks are completed. A different crew type handles different tasks. The tasks are:

1- All areas marked by the DASH report and the Integrity management team are surveyed for leaks. This task is performed by Leak Survey

2- A field service member ensured that all customer reported incidents (dispatch calls) are made safe and relight jobs are completed. This task is performed by Gas Service Reps. (GSRs)

3- All major leaks (grade one) are repaired. This task is performed by Repair Crews

For the purpose of this thesis, we consider the tasks to be independent. This is not necessarily the case since repair crews can not repair a leak before it's found by a leak survey crew or a GSR. However, given that the rate at which leak survey crews find leaks and the number of available GSRs, leak repair crews should have a backlog of jobs to work with during the emergency.

The model is expected to produce the following:

a. Expected restoration finish time for each task (tasks 2 and 3 are pooled together)

b. Crew allocation by division: how much of each crew type a yard gets.

c. Where the crews should come from. For example, San Francisco should get 5 GSRs from San Jose yard and San Jose should get 2 GSRs from Oakland.

The block diagram below summarizes the process, we are modeling the last block in the process:

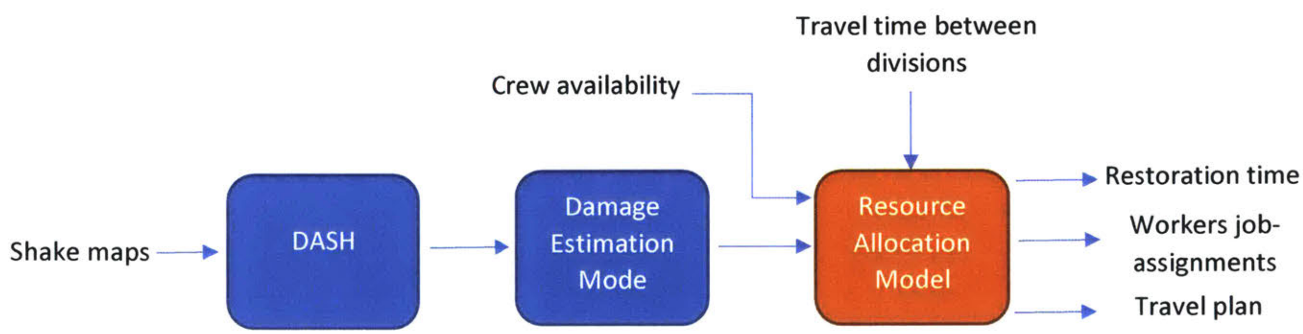


Figure 4-1: Resource allocation model block diagram

4.2 Input Data to the Model

As highlighted in the block diagram (figure 4-1) above, the resource allocation model takes three types of inputs. A summary of the input data for SA 7.9 scenario will be presented in appendix E. In this section, we will describe the inputs to the model. We will describe what data was available to us and what where our assumptions as well.

4.2.1 Worker's availability

We used current worker's shift data to figure out a baseline for the number of worker's available in our simulation. We assumed that 80% of the worker's are be available during the emergency response. Of the available worker's, one third is immediately available and the rest will be available after two-hours of the earthquake. The model can take into consideration that workers are not immediately available to work on restoring the system when allocating the workforce. During a real event, the worker's availability should be constructed based on who is actually available to achieve the optimal outcome.

4.2.2 Expected number of leaks:

The number of leaks were derived from the DASH output as we described in the previous chapter. Using the repair rate for every plat and the length of pipe in that plat, we can aggregate the number of leaks of each type of pipes. We used the pre-generated earthquake scenarios as an input to our model since we did not have earthquakes data to test against. However, given that DASH uses a method that was originally developed for different kind of pipes, this number is treated like an upper limit.

4.2.3 Repair times:

Another input data to our optimization model is the repair times for leaks. The data we had included start and end time for repairs, type of soil (clay, loam, rock, sand, or not-specified) and the type of pipe.

The type of soil did not have significant correlation with the repair times. However, the type of pipe did change the expected repair time. Appendix F have detailed histograms for the repair times. The data is summarized in the table below.

Type of Pipe	Average Repair Time	Standard Deviation	Sample Size
Steel	257 min	279	818
Plastic	284 min	261	775
All	264 min	302	4945

We ran our model with all pipes data since we did not divide the leaks by the pipes type. However, if those data are available, the model does allow us to have different types of leaks and different repair times of each type.

4.3 Problem Formulation

To solve the allocation problem, we modeled the problem as a load balancing work assignment problem. We used two different optimization approaches. The first model has two-stages and has very similar structure to Whipple[10] model. The second model handles combines both stages as one optimization problem.

4.3.1 Baseline Formulation (two-stage optimization)

Stage 1

The goal is to decide how many crews to staff at each location based on the number of estimated repair jobs in each area after an earthquake. Each repair job is assigned to a division.

	Notation	Description
Decisions	X_{jk}	Binary variable that indicates if job j is assigned to yard k
	C_k	Variable representing yard k workload
	C_k^*	Number of crews assigned to yard k
	C	Variable representing total workload
Data	y_{jk}	Time it takes to finish job j from yard k
	M_k	Maximum capacity of yard k
	m_k	Minimum number of crews required to stay in yard k
	C^*	Total number of crews

The mathematical model is represented as the following MIP problem:

Objective Minimize C

S.T.

$$\sum_k C_k \leq C \qquad (4\text{-}1)$$

$$\sum_j y_{jk} X_{jk} \leq C_k \qquad (4\text{-}2)$$

$$\sum_j X_{jk} = 1 \qquad (4\text{-}3)$$

$$C_k^* \leq M_k \qquad (4\text{-}4)$$

$$C_k^* \geq m_k \qquad (4\text{-}5)$$

$$X_{jk} \in \{0,1\} \qquad (4\text{-}6)$$

Where C_k^* is defined as

$$C_k^* = \frac{C_k}{C} \times C^* \qquad (4\text{-}7)$$

Constraint (4-1) ensures that no workload C_k exceeds the worst-case load C. Constraint (4-3) ensures that all jobs are attended. Division capacities and minimum staffing are ensured by constraints (4-4) and (4-5).

Since division cover a large area, we have decided to move away from trying to assign a job to division and instead assume a job will be handled by the division where the leak is estimated to be.

Stage 2

Now that we've decided how many repair crews to station at each division, we can determine how to reach that allocation. This problem can formulated as a network min-cost flow problem.

	Notation	Description
Decisions	$f(v,w)$	Variable that indicates the number of worker's between divisions $v\ and\ w$
Data	$G(V,E)$	G = (V,E) where V represents divisions and E represents travel routes
	$c(v,w)$	Travel costs in hours between two divisions
	$b(v)$	Number of crews required (or available) at a division
	$u(v,w)$	Road capacity between two divisions

$b(v)$ is calculated as the difference between the current staffing at a division and what it needs according to the solution of the first formulation

$$\min \sum_{(v,w)\in E} c(v,w)\, f(v,w)$$

subject to

$$\sum_{w\in V} f(v,w) - \sum_{w\in V} f(w,v) \le b(v)\ \forall\ v \in V \qquad (4\text{-}1)$$

$$f(v,w) \ge 0\ \forall\ (v,w) \in E \qquad (4\text{-}2)$$

$$f(v,w) \le u(v,w)\ \forall\ (v,w) \in E \qquad (4\text{-}3)$$

After developing the previous formulation and discussing it with PG&E, it didn't make sense to look at assigning jobs to divisions for multiple reasons. The main reason is that most jobs are automatically assigned to the division where the leak falls under. Moreover, supervisors operating from the basecamps are more familiar with their area and assigning having supervisors from different divisions manage the jobs could add unwanted delays.

Thus, we can determine the best allocation (stage one) by calculating the workload ratio of every division relative to the total workload and assigning workers based on the relative workload ratio.

More specifically,

$$C_k^* = \frac{C_k}{C} \times C^*$$

$$C_k = j_k$$

$$C = \sum_K C_k$$

Where j_k is the number of leaks located in division *k.*

4.3.2 Combining Crew Allocation with Travel (1-stage optimization)

In the previews section, we formulated the problem by decoupling the allocation problem and the crew routing problem. This could lead to a suboptimal solution. First, it lacks the ability to represent that a worker can work in multiple divisions. For example, a worker can start by attending the work in their initial division and then go to a different division once they are done. Moreover, it artificially increases the available hours by not taking into consideration the wasted travel time.

To model those limitations, we consider traveling between locations as a special job that adds to the workload of the crew. In addition, unlike the previous formulation, we look at the problem from a crew standpoint. Our decision variable indicates whether a job is a assigned to a specific crew or not. In contrast, we assigned jobs to divisions in the previous model.

In this model, we create another decision variable, which indicates which workers need to travel to a specific location. Workers can not work on a job in division ***k*** unless they are based in that location or they travel there. A sided benefit of this formulation is that we can also assign dummy jobs to crews who are not going to be available at the beginning of the restoration effort but are expected to show up few hours after the work had started.

In this formulation, the restoration time is the time it takes the last crew to finish their job. In its essence, it is a load-balancing problem between crews. In the figure below, the restoration time would be 24 hours.

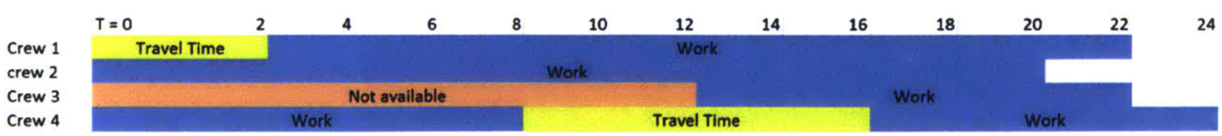


Figure 4-2: Sample output from the resource allocation model for each worker

	Notation	Description
Decisions	X_{ijk}	Binary variable that indicates if worker i is assigned to job j in yard k
	Y_{ik}	Binary Variable that indicates if worker i travels from home location to yard k
	L	Variable representing the maximum workload across all workers in minutes
Data	t_{ik}	time it takes worker i to travel to yard k
	d_k	Average time of a job in yard k
	h_{ik}	Parameter that indicates iif worker i home location is yard k

LP formulation:

Min L

s.t.

$\sum_j X_{ijk} d_j + \sum_k Y_{ik} t_{ik} \leq L$ for all i ∈ W (4-1)

$X_{ijk} \leq h_{ik} + Y_{ik}$ for all i,j,k ∈ W,J,K (4-2)

$\sum_k Y_{ik} \leq 1$ for all i ∈ W (4-3)

$\sum_i \sum_k X_{ijk} \geq 1$ for all j ∈ J (4-4)

$X_{ijk}, Y_{ik} \in \{0,1\}$ for all i,j,k ∈ W,J,K (4-4)

Constraint (4-2) ensures that workers can only do work in their home division or in another division if they take the trip to that division. Constraint (4-3) limits workers travel to one trip during the planning period. We impose this constraint for both operation and simplification reasons. Constraint (4-4) ensures that all jobs are attended. The worst case workload is enforced by constraint (4-1)

One issue with this formulation the size of the MIP. We have $O(i \times j \times k)$ decisions variables. This leads to a very long run time when trying to solve the problem to optimality. The solver took more than a day to reach a 5% gap, which is considered long time for the intended purpose. Even if kept running for another day, it never found the optimal solution.

Since the previous formulation was taking a long time, we decided to change reduce the size of the search space. The data we have only includes average repair times and doesn't indicate how long a specific job is expected to take. Thus, we modified the formulation slightly to reduce the number of variables. Instead of having a variable for every crew for every job, we have a variable that indicates the number of jobs assigned to a crew in a specific location. We also added another term that indicates if a worker won't be available at the begging of the restoration efforts as suggested by the company. The modified formulation:

	Notation	Description
Decisions	X_{ik}	Integer variable representing the number of jobs assigned to worker i in location k
	Y_{ik}	Binary Variable that indicates if worker i travels from home location to yard k
	L	Variable representing the maximum workload across all workers in minutes
Data	t_{ik}	time it takes worker i to travel to yard k
	d_k	Average time of a job in yard k
	h_{ik}	Parameter that indicates iif worker i home location is yard k
	a_i	Expected unavailability of worker i
	j_k	Expected number of jobs in yard k

$Min\ L$

S.T

$\sum_k X_{ik} d_k + \sum_k Y_{ik} t_{ik} + a_i \leq L$ for all i ∈ W (1)

$X_{ik} \leq 200\ (h_{ik} + Y_{ik})$ for all i,k ∈ W,K (2)

$\sum_k Y_{ik} \leq 1$ for all i ∈ W (3)

$\sum_i X_{ik} \geq j_k$ for all k ∈ K (4)

$Y_{ik} \in \{0,1\}$ for all i,j,k ∈ W,J,K (5)

$X_{ik} \in \mathbb{Z}$ for all i,j,k ∈ W,J,K (4-4)

(1): worst case workload is the sum of jobs assigned to a worker, the travel time, and their initial unavailability

(2): Workers can not be assigned to jobs unless it's in their home location or they travel. 200 is an arbitrary large number that we do not expect we will reach. It is just an upper limit of how many works can be assigned to a worker in one location.

(3): A worker can only travel to one other yard

(4): All jobs in a division must be attended

4.4 Solving the Problem with Uncertainty

The models we presented in the previews section assumes that all the data is available before we run the model. However, this is clearly not the case especially when running the model during the first few hours of the restoration efforts. There are three sources of uncertainty in the model. Those sources are:

1. **The number of required repairs at each division**: The number of damages in each division is based on the damage model. As mentioned in chapter x, this model was tested only on small-scale earthquake. However, based on literature review and the small-scale cases PG&E faced, I can confidently say that this model is close to a worst-case scenario [PGA paepr]. This mean that by solving for this scenario, our solution will be valid for almost all scenarios.
2. **Crew availability**: The model right only takes into account available crews or crews expected to arrive at a certain point in the future. However, if the crews are delayed or they don't show up, the solution won't be valid. However, given that the optimization solver runs relatively fast, it's possible to run the model again if new information regarding crew availability becomes available
3. **Repair times:** In our optimization, we assumed all repairs take a nominal value d_k, However, this will never be the case.

It is important to program those uncertainties into the model. More specifically, we want to ensure our solutions are always valid despite the variation in repair-time values. Recall from our previous relaxed model the only constraint using unknown workload (d_k) was the following:

$\sum_k X_{ik} d_k + \sum_k Y_{ik} t_{ik} + a_i \leq L$ for all i ∈ W (1)

To consider all possible value of (d_k), we can change this constraint and ensure our optimization is still valid for all values of (d_k)

$\sum_k X_{ik} d_k + \sum_k Y_{ik} t_{ik} + a_i \leq L \quad \forall\, i \in W,\ \forall\, d_k \in U$ (1)

We define $U$to be the set of all values $d_k$can potentially take on. This new constraint ensures that all solutions from this model are valid. However, the model is no longer a MIP. We can re-model the formulation back to a MIP using a box constraint.

4.4.1 Robust Optimization Using a Box Constraint

To reformulate the program as an MIP program again, a simple way is to use box constraint. We simply assume that all d_{ik} values fall in the range $d_{ik} \in [l_{ik}, u_{ik}]$. Now our uncertainty set is the following:

$$U_k = \{d | \forall i, l_{ik} \leq d_{ik} \leq u_{ik}\}$$

Now our constraint from the previous section:

$$\sum_k X_{ik} d_k + \sum_k Y_{ik} t_{ik} + a_i \leq L \qquad \forall \text{ i} \in \text{W}, \forall d_k \in U \qquad (1)$$

can be rewritten as follows:

$$\sum_k X_{ik} u_{ik} + \sum_k Y_{ik} t_{ik} + a_i \leq L \qquad \forall \text{ i} \in \text{W} \qquad (1)$$

In this formulation, we are assuming all repairs take the worst case scenario. This ensures that our solution is valid and easy to implement. However, given that repair times are likely to be i.i.d, it's extremely unlikely that all the repairs take the worst case value which lead to a significantly higher and conservative objective value [8]. In the next section, we present a method to balance robustness and performance using Bertsimas-Sim uncertainty sets which they presented in their paper "The Price of Robustness" [8]

4.4.2 Robust Optimization Using Bertsimas-Sim Uncertainty Sets

Bertsimas and Sim [8] proposed an approach that balances robustness with performance. They present a flexible formulation that offers a way to trade between robustness and performance.

Let d_{ik} take values in a distribution in the interval $[d_{ik} - \widehat{d_{ik}}, \ d_{ik} + \widehat{d_{ik}}]$. We introduce parameter Γ that indicates the number of leaks that take the worst case. The role of Γ is to adjust the robustness of the solution. The goal is to protect against all cases up to Γ. We will discuss the ideal selection for Γ in the next section.

We can represent this uncertainty set as follows:

$$U_k = \{d | \forall i, d_{ik} \in \left[d_{ik} - \widehat{d_{ik}}, d_{ik} + \widehat{d_{ik}}\right], \sum_i \frac{|d_{ik} - \overline{d_{ik}}|}{d_{ik}} \leq \Gamma\}$$

We can make the constraint robust by dictating the following

$$d_{ik} = \bar{d}_{ik} + \widehat{d_{ik}}\ \hat{u}_{ik}$$

Where $\bar{d}_{ik}$ is the nominal repair time value and $\hat{d}_{ik}$ is the potential deviation from the nominal value. The total deviation from the nominal value must be bounded for all $\widehat{d_{ik}}$ and therefore all $\hat{u}_{ik}$ must reside in the following uncertainty set:

$$U_{k,u} = \{u|\forall i, u_{ik} \in [-1,1]; \sum_i |u_{ik}| \leq \Gamma\}$$

Using this representation, the original constraint can now be rewritten as follows:

$$\textstyle\sum_k X_{ik} d_{ik} + \sum_k Y_{ik} t_{ik} + a_i + \max\limits_{u \in U_{k,u}} \sum_i X_{ik}\, \hat{d}_{ik}\, u_{ik} \leq L \qquad \forall\ \mathrm{i} \in \mathrm{W} \qquad (1)$$

The max problem on the right hand side is a liner optimization problem and can be formulated as follows:

Maximize $\sum_i X_{ik}\, \hat{d}_{ik}\, u_{ik}$

subject to

$\sum_i u_{ik} \leq \Gamma$ for all *k*

$0 \leq u_{ik} \leq 1$ for all u ∈ *U*

Following Bertsimas-Sim model, we can show that the problem can be re-formulated as follows:

Min L

S.T.

$\sum_k X_{ik} d_{ik} + \sum_k Y_{ik} t_{ik} + u_i + z_i \Gamma_i + \sum_k p_{ik} \leq L_i$ for all i ∈ W (1)

$X_{ik} \leq 200\ x\ (h_{ik} + Y_{ik})$ for all i,k ∈ W,K (2)

$\sum_k Y_{ik} \leq 1$ for all i ∈ W (3)

$\sum_i X_{ik} \geq j_k$ for all k ∈ K (4)

$L \leq L_i$ for all i ∈ W

$z_i + p_{ik} \geq \hat{d}\, s_{ik}$ for all i,k ∈ W,K

$x_{ik} \leq s_{ik}$ for all i,k ∈ W,K

$p_{ik} \geq 0$ for all i,k ∈ W,K

$z_i \geq 0$ for all i ∈ W

$s_{ik} \geq 0$ for all i,k ∈ W,K

This new formulation have a robust solution. The solution still can be infeasible for some values distributions of repair times. Larger Γ values produce more robust solutions but it will potentially reduce the optimality of the solution.

In the next section, we will discuss how to choose the values for Γ and d_{ik} by examining the historical repair times distributions and noting how that led to our decisions for Γ and d_{ik}.

4.4.3 Determining the Robustness of Solution

Determining the ideal parameters values for the robust optimization requires that we look at and analyze the historical repair data. Fortunately, PG&E has the information about all leak repairs all the way back to 2009. Using this data, we can try to look at the repair time distribution to determine the ideal parameters for the robust optimization. The distribution of the leaks' repair time looks as follows:

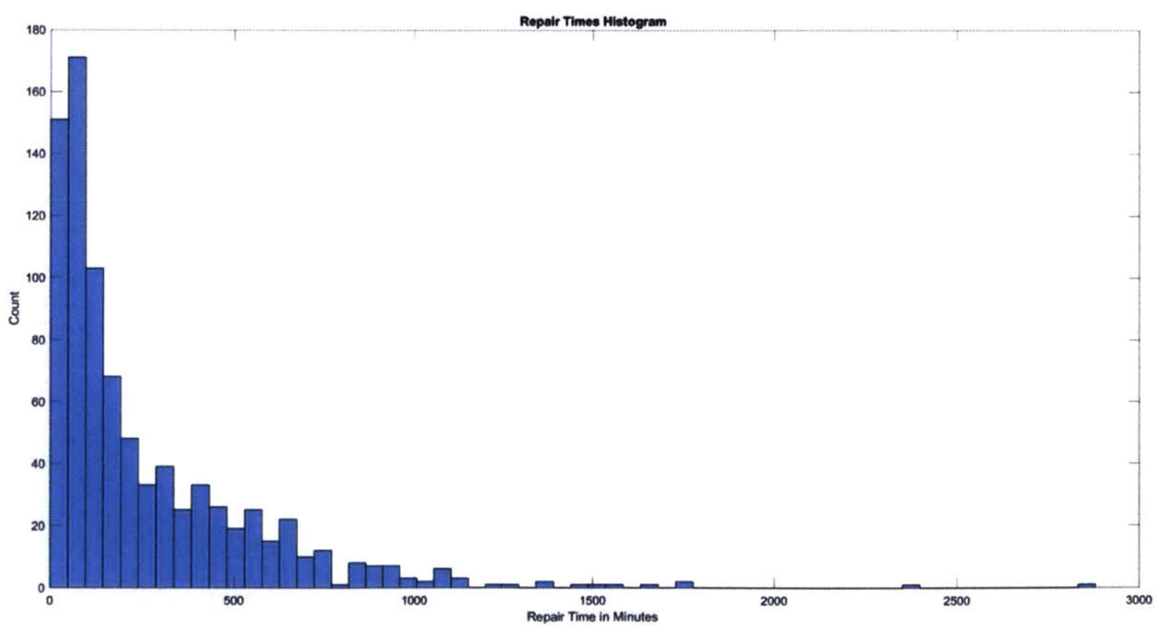


Figure 4-3: Repair times distribution histogram

After excluding some outliers, we choose the nominal d_{ik} to be the average and the half width to be 2 times the standard deviation. This leads to $\overline{d_{ik}} = 264.61$ and $\widehat{d_{ik}}$ = 604.1.

To choose Γ, we ransimulations for each scenario by selecting random draws with replacements from the repair times' histograms. Then, we can calculate the relative error from the average repair time as follows:

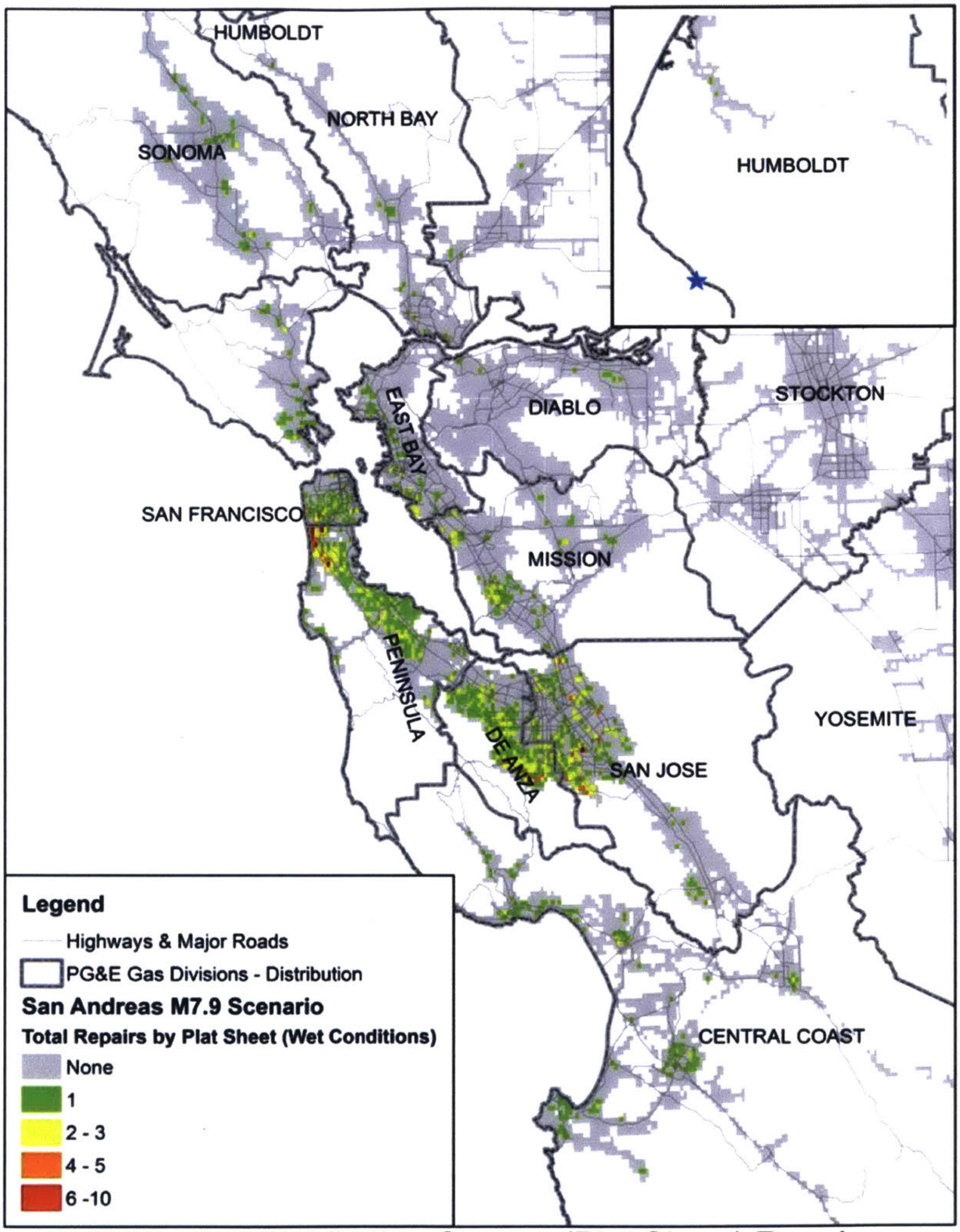
$$\Gamma = \sum_i \frac{|d_{ik} - \overline{d_{ik}}|}{d_{ik}}$$

We ran 10,000 simulations. For each simulation, we used the average number of repairs per worker with generated using the model from the previous section under different crew availability scenarios and earthquake scenarios.

To ensure that our solution will be valid with a high probability, we selected Γ such that in incorporates a significant portion of the histogram. We believe a 95th percentile is sufficiently robust to meet the operations needs of PG&E. Under the San Andreas 7.9 magnitude earthquake (SA7.9), the 95th percentile $\Gamma = 6.18$ under the assumption that all crews will be available. Appendix B contains the simulation results of different scenarios.

4.5 Comparison of Optimization Results

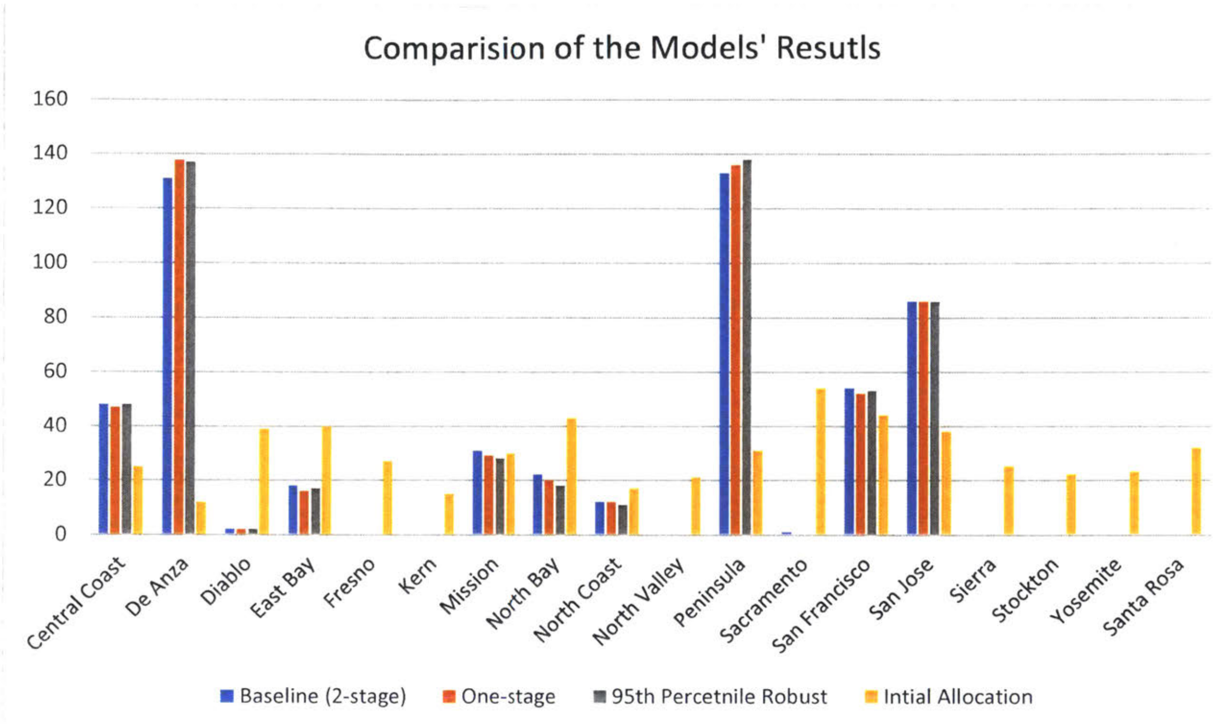
To compare the formulations, we tested them on the SA7.9 scenario. We started by generating the predicted damages at each division using the shake maps from USGS and the DASH models. The graph and tables below summarize how the damages are distributed.



Estimated Distribution System (Plat Sheet) Repairs

Location Name	Repairs	Location Name
Central Coast	496	Central Coast
De Anza	1367	De Anza
Diablo	22	Diablo
East Bay	192	East Bay
Fresno	0	Fresno
Kern	0	Kern
Mission	319	Mission
North Bay	225	North Bay
North Coast	129	North Coast
North Valley	0	North Valley
Peninsula	1384	Peninsula
Sacramento	8	Sacramento
San Francisco	563	San Francisco
San Jose	891	San Jose
Sierra	0	Sierra
Stockton	0	Stockton
Yosemite	0	Yosemite
Santa Rosa	0	Santa Rosa

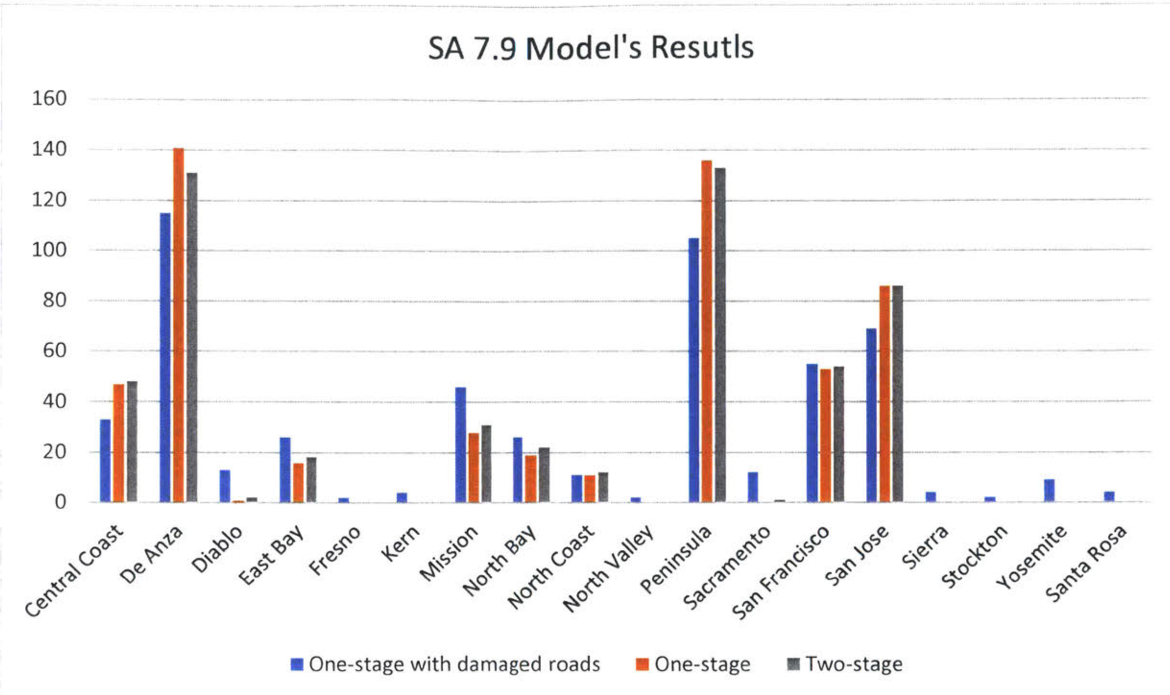
Appendix D has a map with that shows how where from and to workers traveled. The results of each of the formulation is summarized in the figure below:



All three models produce very similar results and within margin of error of the data. The results indicates that the base model allocates more crews to area's that are not the most affected areas by the earthquake. This happens because in that formulation, we do not consider the travel times when deciding on the ideal allocation. This artificially increase the time crews who travel to more damaged area have. Aside from that, there does not appear to be any difference in allocation between the robust formulation and the deterministic version of the formulation.

We wanted to carry out further investigation on how and when the models could differ; we took into consideration the changes in the network due to the earthquake. More specifically, we wanted to see how the effect of inaccessible roads and bridges would affect the results of the two-stage optimization and the one-stage optimization results.

To get the results, we ran the models again but we modified the time it takes to get through highways by a factor of two. We also made all bridges in the affected area inaccessible. We got the following results:



We can see in the data that running the model with increased travel times doesn't move crews around as much as the other two runs. If we look at De Anza, and the Peninsula divisions, we can see 20%-40% reduction in the allocation relative to the other two cases. The results suggest that it's not always better to send crews to the most damaged area's if a significant portion of their time will be wasted in travel.

Tying it all together, if we do not consider travel times when allocating (two-stage optimization), we might not allocate enough worker's to the most damaged area. This will lead to a sub-optimal solution. However, if the roads are severely damaged, it might not make sense to send as many workers' to the area's that are severely damaged since the time they spend traveling could be utilized somewhere else in the system. Assuming that travel conditions improve after the first few hours, PG&E planners might consider moving the workers to help in areas that are close to the most damaged divisions. Once road conditions are improved, they can run another optimization to determine if they need to send more workers to help. This is extremely relevant in cases where there are major accessibility issues. For example, the Peninsula area is mainly accessible through bridges that are not going to be available after a major earthquake.

Another approach PG&E can look into is to fly workers into the most damaged area. Such a decision should be made when the expected restoration time is too long and many workers are not being utilized since they can not reach their ideal allocation dentation.

4.6 Conclusion

We presented a model that be used a guideline for PG&E to allocation their crews after an earthquake hits. In this model, we accounted for errors. The model should produce a robust solution that is sufficiently optimal. This will help in improving the restoration time of the over a completely manual allocation

In our model, we did not consider all the factors that PG&E uses to make their decision. For example, if possible, PG&E prefers to keep their crews in their home location. In theory, the optimal solution would be to send worker from division A to division B and send another worker from division B to division C at the same time. However, this leads to two workers leaving their home division. In such a case, PG&E would opt to send a worker directly from division A to C and not send anyone from Division B to C. Some other factors we didn't consider is the potential risk of the leaks. All leaks we consider are classified as 1 (must be fixed as soon as possible). However, PG&E could send more workers to areas with less leaks because they have more important customers in that area (e.g. hospitals). Given that not all factors are present, PG&E is unlikely to follow the results for the model exactly but it can provide a concrete sense on how should PG&E allocate their crews

5 Conclusions and Future Work

During the initial stages of the projects, we identified multiple directions where we could have improved upon in the current process. We have explored some of those areas while we have not talked about other areas that could affect the restoration effort. In this chapter, we will talk about some of the future work we identified that PG&E should explore in the future.

5.1 Improvements to the damage prediction model:

As mentioned in chapter 3, we did not have enough data at PG&E to further improve dash or create a different model that could supplement it. We tried to look at earthquake data from different utilities but given the limited time of the project, we did not think it was feasible.

PG&E can leverage the data available from earthquakes like Fukushima in Japan or the more recent August 2016 earthquake in Italy. Even though the infrastructure is likely to be different, it will improve the damage output currently generated by DASH.

In addition, PG&E should establish a mechanism to capture data related to earthquakes. More specifically, if an earthquake hits in the future, we recommend that PG&E tags leaks, dispatch calls and other types of damages related to earthquakes. This would allow PG&E to perform analysis on this data to improve their damage prediction models in the future.

5.2 Improvements to the resource allocation model:

5.2.1 Decreasing the Uncertainties in Repair Times

Once we have a better granularity of locating damages, we can use repair times that for the specific types of leaks we expect. In this project, we explored the factors that contribute to repair times such as the type and size of pipes. Some of those factors were statistically significant. However, our damage prediction model lacked the specificity to apply those differences to the leaks we expect to happen. So we had to opt out and use the nominal repair times instead.

We believe that further exploring the repair times data combined with a better damage prediction model could significantly reduce the uncertainty of the model and lead to better allocation of resources.

5.2.2 Examining the Relation between Crew Availability and Earthquakes

The availability of the crews is currently an input to the model. However, by knowing the work location of the crews, we can estimate which crews are not going to be available. For example, we could predict that workers in the affected area won't be available immediately. W We can potentially estimate the percentage of crews who are not going to be available. This is useful because PG&E currently does not have a mechanism to verify who will be available other than the crews calling-in after an earthquake hits to confirm their availability. This will help PG&E in coming up with an allocation plan as soon as an earthquake hits.

5.2.3 Increasing the Granularity of the Model

Right now, the workers are assigned to divisions. We have decided to take this route given that the damage data we have are not granular enough. A model to allocate workers to different yards/base camps within division based on travel times could further improve the efficiency of the restoration efforts.

5.3 Other Potential Improvements

5.3.1 Studying the Availability of Base Camps Space and Equipment

Currently PG&E have agreements with different entities to use their facilities as base camps during emergencies. The location and capacity of basecamps will be decided during the emergency. While the basecamps capacity is sufficient in theory. In practice, those base camps are shared with different entities and are very likely not going to provide sufficient space to host all crews in one location. Similarly, PG&E have contracts to get extra equipment to build the basecamps during emergencies. The companies that provide the equipment have contracts with most other emergency responders in the area (e.g. fire department). In the past, there were not enough equipment to meet the demand and whoever requested the equipment first got it. Those limitations could affect the allocation model results and should be further explored in the future.

5.4 Conclusion:

Earthquakes damage to utilities is unavoidable. Proper planning and execution during and after the event ensures that assets are repaired in a timely manner to ensure the public's safety and reduce the impact on communities.

In this paper, we provided data driven approach to help PG&E improve their earthquake emergency response planning. Those models will continue to develop as more data becomes available and as part of the earthquake preparation efforts that this project was part of. Part of the effort was running simulations on potential earthquake. PG&E can use the output of those simulations during earthquake emergencies if actual data is not available due to service interruption from earthquakes.

The model's we developed only looked at the gas business line of PG&E. They can be used in different service lines and under different types of emergencies as part of PG&E's effort for continuous improvement.

References:

[1] American Lifelines Alliance, ALA (2001). Seismic Fragility Formulations for Water Systems, American Society of Civil Engineers (ASCE) and Federal Emergency Management

[2] Seismic Damage Estimation for Buried Pipelines: Challenges after Three Decades of Progress Omar Pineda-Porras-Mohammad Najafi - Journal of Pipeline Systems Engineering and Practice - 2010

[3] FEMA. 1999. Earthquake loss estimation methodology HAZUS-MH— Technical manual, http://www.fema.gov/hazus 2001.

[4] Ming-Jong Yao, K Jo Min, Repair-unit location models for power failures, IEEE Transactions on Engineering Management, 45:57-65, 1998.

[5] Sudipto Guha, Anna Moss, Joseph Seffi Naor, Baruch Schieber. Efficient recovery from power outage. Conference proceedings of the annual ACM symposium on theory of computing 199-. P. 574-92, 1999.

[6] Willy Herroelen, Roel Leus. Project scheduling under uncertainty: Survey and research potentials: European Journal of Operational Research, 165(1):289-306, 2005.

[7] Siddarth Balwani, Operational Efficiency through Resource Planning Optimization and Work Process Improvement. Master's thesis Massachusetts Institute of Technology, 2012.

[8] Bertsimas, D., & Sim, M. (2004). The Price of Robustness. Operations Research, 52(1), 35-53. doi:10.1287/opre.1030.0065

[9] Bertsimas, D. & Sim, M. Math. Program., Ser. B (2003) 98: 49. Robust discrete optimization and network flows doi:10.1007/s10107-003-0396-4

[10] Sean Whipple. Predictive Storm Damage Modeling and Optimizing Crew Response to Improve Storm Response Operations Master's thesis, Massachusetts Institute of Technology, 2014.

[11] Fiedrich, F., Gehbauer, F., & Rickers, U. (2000). Optimized resource allocation for emergency response after earthquake disasters. Safety Science, 35(1-3), 41-57. doi:10.1016/s0925-7535(00)00021-7

Appendix A: Dispatch Calls Regression Models:

(1) Simple Regression Model:

```
Linear regression model:
    DispatchCalls ~ 1 + LND_SLIDE + CORRISVITY + Number_Of_Cus + PGV

Estimated Coefficients:
```

	Estimate	SE	tStat	pValue
(Intercept)	-0.93391	0.21464	-4.3512	1.4454e-05
LND_SLIDE	-0.031314	0.03706	-0.84496	0.39827
CORRISVITY	0.0056496	0.055687	0.10145	0.9192
Number_Of_Cus	0.007175	0.00042551	16.862	1.4232e-58
PGV	0.050087	0.0024637	20.33	2.1917e-81

```
Number of observations: 1511, Error degrees of freedom: 1506
Root Mean Squared Error: 1.75
R-squared: 0.326,  Adjusted R-Squared 0.324
F-statistic vs. constant model: 182, p-value = 3.42e-127
```

(2) Pair-wise Interactions:

```
Number of observations: 1511, Error degrees of freedom: 1506
Root Mean Squared Error: 1.75
R-squared: 0.326,  Adjusted R-Squared 0.324
F-statistic vs. constant model: 182, p-value = 3.42e-127
>> mdl = stepwiselm(train,'interactions')
1. Removing LND_SLIDE:Number_Of_Cus, FStat = 0.015853, pValue = 0.89982
2. Removing LND_SLIDE:CORRISVITY, FStat = 0.1442, pValue = 0.70419
3. Removing LND_SLIDE:PGV, FStat = 0.52634, pValue = 0.46826
4. Removing CORRISVITY:Number_Of_Cus, FStat = 1.9659, pValue = 0.16109

mdl =


Linear regression model:
    DispatchCalls ~ 1 + LND_SLIDE + CORRISVITY*PGV + Number_Of_Cus*PGV

Estimated Coefficients:
```

	Estimate	SE	tStat	pValue
(Intercept)	-0.80312	0.30648	-2.6205	0.0088682
LND_SLIDE	0.069426	0.032602	2.1295	0.033378
CORRISVITY	0.23182	0.098314	2.358	0.018503
Number_Of_Cus	-0.0025837	0.00056008	-4.6131	4.3043e-06
PGV	0.048099	0.023644	2.0343	0.042095
CORRISVITY:PGV	-0.016831	0.0076699	-2.1945	0.028354
Number_Of_Cus:PGV	0.00045192	1.9693e-05	22.949	3.5945e-100

```
Number of observations: 1511, Error degrees of freedom: 1504
Root Mean Squared Error: 1.5
R-squared: 0.502,  Adjusted R-Squared 0.5
F-statistic vs. constant model: 253, p-value = 1.34e-223
```

Appendix B: Robustness Parameter "Γ" Simulation Results:

Scenario	SA 7.9		50th percentile	90th percentile	95th percentile	98th percentile
#Available Workers		500	4.05	5.64	6.18	6.84
		250	8.25	10.4	11.09	11.95
		100	20.89	24.16	25.21	26.37

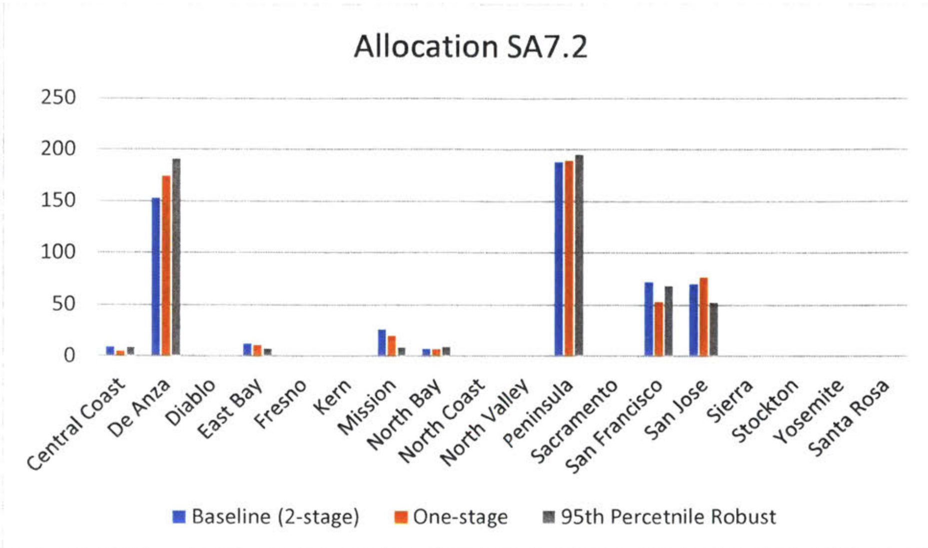
Scenario	HW 7.0		50th percentile	90th percentile	95th percentile	98th percentile
#Available Workers		500	1.59	2.67	3.15	3.65
		250	3.34	4.81	5.32	5.94
		100	8.94	11.18	11.91	12.77

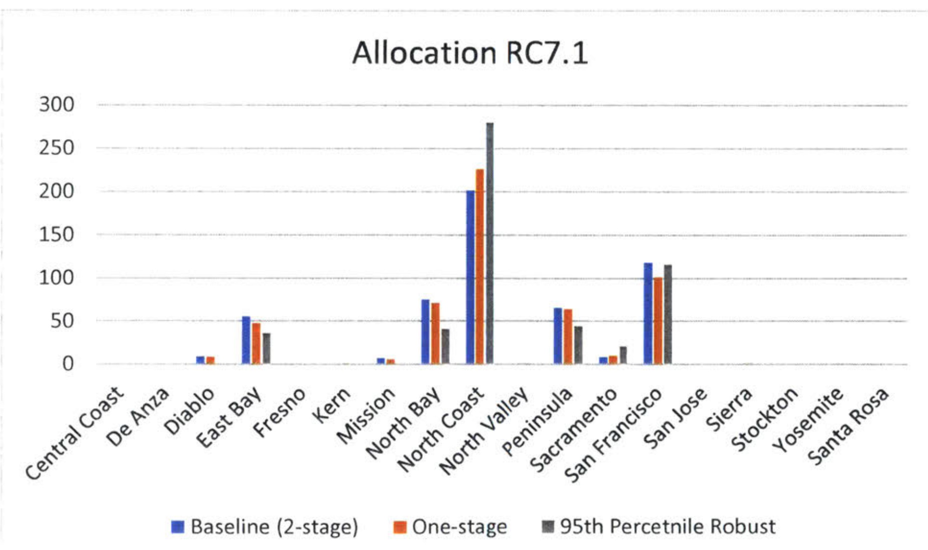
Scenario	RC 7.1		50th percentile	90th percentile	95th percentile	98th percentile
#Available Workers		500	0.61	1.21	1.55	2.08
		250	1.25	2.20	2.65	3.16
		100	3.34	4.81	5.34	5.95

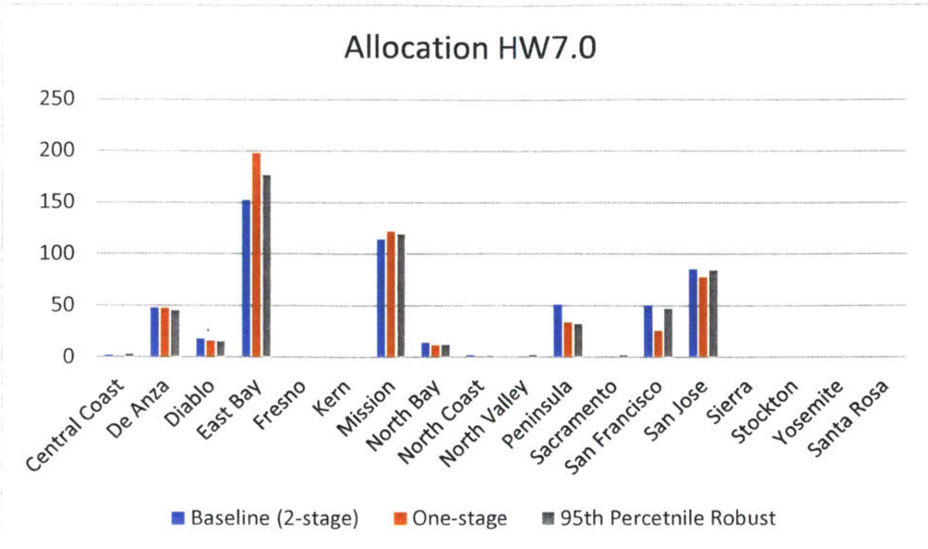
Scenario	SA 7.2		50th percentile	90th percentile	95th percentile	98th percentile
#Available Workers		500	1.94	3.11	3.59	4.11
		250	4.38	6.04	6.59	7.27
		100	11.40	13.89	14.69	15.64

Scenario	Napa 6.0		50th percentile	90th percentile	95th percentile	98th percentile
#Available Workers		500	0.00	0.00	0.00	0.00
		250	0.00	0.00	0.00	0.00
		100	0.31	0.63	0.96	1.37

Appendix C: Crew Allocation Results for Different Scenarios:

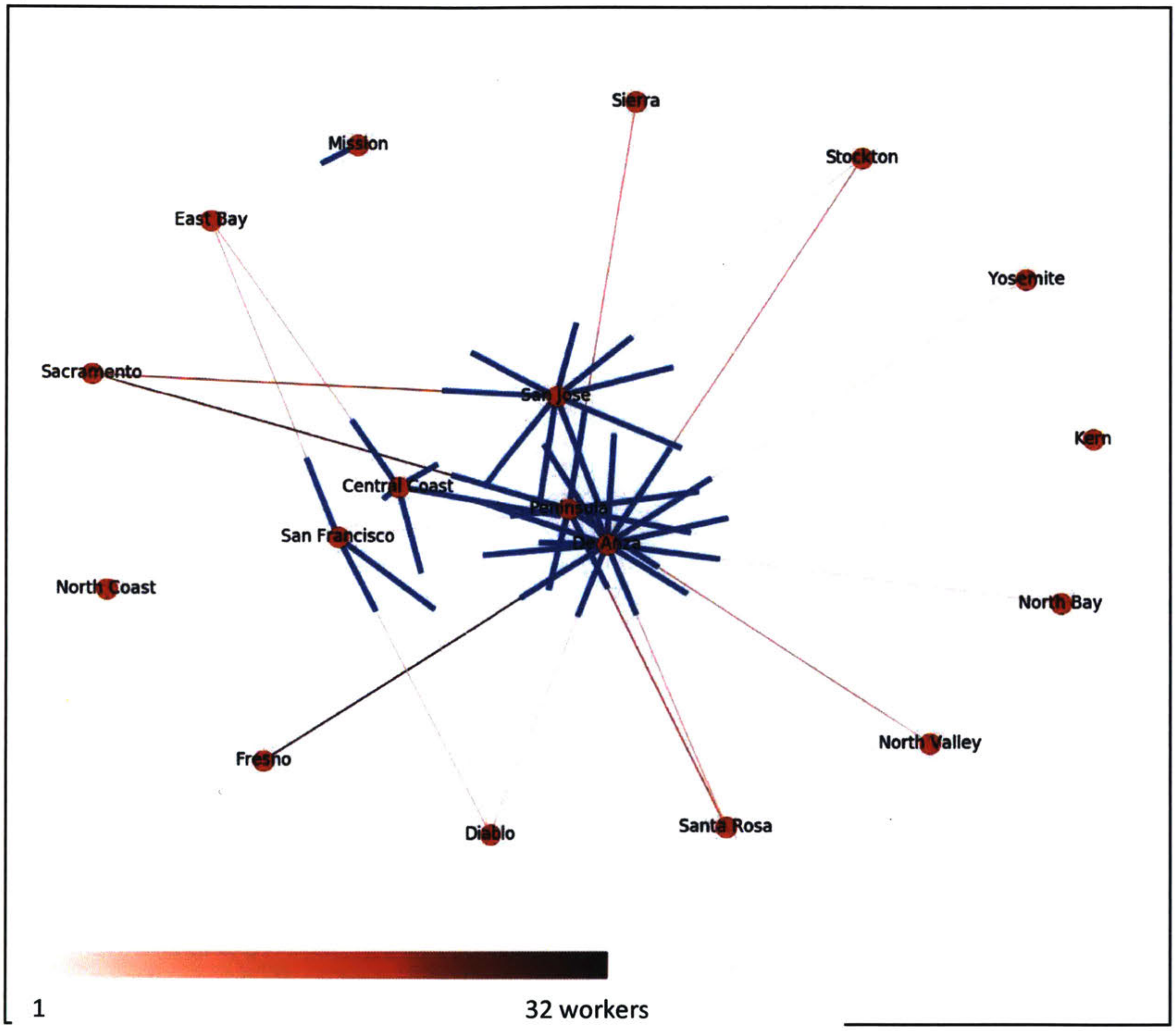





Allocation HW7.0
250
200
150
100
50
0
Central Coast
De Anza
Diablo
East Bay
Fresno
Kern
Mission
North Bay
North Coast
North Valley
Peninsula
Sacramento
San Francisco
San Jose
Sierra
Stockton
Yosemite
Santa Rosa
Baseline (2-stage)
One-stage
95th Percetnile Robust

Appendix D: SA7.9 Travel Summery Map

The network graph below shows the number of workers travelling between divisions from running the optimization model.



Appendix E: Data used to run the SA7.9 earthquake scenario optimization

1: Worker's Availability: the following table shows the aggregate available repair workforce used for the model. Note that individual workers had different availability (hours/day) and initial unavailability (how long before they start)

Location Name	Available Worker's During Emergencies
Central Coast	25
De Anza	12
Diablo	39
East Bay	40
Fresno	27
Kern	15
Mission	30
North Bay	43
North Coast	17
North Valley	21
Peninsula	31
Sacramento	54
San Francisco	44
San Jose	38
Sierra	25
Stockton	22
Yosemite	23
Santa Rosa	32

2: Travel Times: the data represents the travel time in minutes from division (row) to division (col)

From/To	Central Coa	De Anza	Diablo	East Bay	Fresno	Kern	Mission	North Bay	North Coas	North Valle	Peninsula	Sacrament	San Francis	San Jose	Sierra	Stockton	Yosemite	Santa Rosa
Central Coa	1	30.5	82.6	70.2	157	345	46.8	115	242	276	68	147	73.3	32.1	178	107	116	129
De Anza	30.5	1	64.8	48.4	158	313	25	97	248	258	38.5	129	43.5	10.4	161	89.1	98.2	97.1
Diablo	82.2	64.4	1	22	176	279	40.8	35.8	271	197	38	67.6	30.8	55.4	99.4	51.7	80.3	62.2
East Bay	70.5	48.5	22.6	1	177	275	26.5	43.2	272	211	19.5	81.8	12.4	40.6	114	71.6	80.7	58.4
Fresno	158	158	177	177	1	451	167	209	109	333	196	173	188	152	201	127	96.8	234
Kern	345	313	280	275	450	1	301	258	546	147	275	289	271	315	318	325	355	217
Mission	47.4	25.4	41	26.3	170	301	1	73.2	257	235	39.4	105	38	17.5	137	65.3	74.4	84.7
North Bay	113	95.3	34.9	42.3	207	256	68.8	1	302	190	55.5	60	48.3	86.3	91.8	70.8	111	39.8
North Coas	242	247	272	272	109	545	257	304	1	446	285	285	283	241	307	233	202	329
North Valle	276	258	198	211	333	147	234	190	446	1	224	162	217	249	190	209	237	227
Peninsula	67.9	38.5	38.4	19.2	195	275	39.3	56.2	285	224	1	94.8	9.5	47.4	127	89.9	99.1	58.6
Sacrament	147	129	68.6	81.9	171	289	106	61.5	286	162	95.1	1	87.9	120	33.2	48.8	75.3	97.6
San Francis	73.1	43.4	31.4	12.3	188	271	38	49.3	283	217	9.6	87.9	1	48.3	120	83	92.1	54.6
San Jose	32.7	10.7	55.9	40.7	152	316	17.3	88.2	241	250	47.5	120	48.4	1	152	80.3	89.4	99.2
Sierra	179	161	100	114	201	317	137	93.2	307	191	127	32.9	120	152	1	80.4	106	129
Stockton	108	89.7	50.8	71.6	127	326	66.1	73	239	210	90.3	49.2	83.2	80.7	83.9	1	31.7	109
Yosemite	117	98.8	80.8	80.7	96.8	355	75.2	113	202	237	99.4	76.4	92.3	89.8	105	31.3	1	138
Santa Rosa	129	97.1	63.3	58.5	234	217	85	41.4	329	227	58.9	97.8	54.7	99.1	130	109	138	1

3: Leaks Data: Summarized in section 4.5