

Optimizing Ambulances in Vance County

Ayush Batra, Woonggyu Jin, Sophia Yang

October 26, 2025

1 Background and Motivation

Emergency medical services (EMS) are a critical part of the medical system, responsible for saving lives in times of crisis. Common medical situations EMS are responsible for include emergencies such as cardiac arrest, penetrating trauma, brain injury, stroke, and overdose. The faster EMS can reach a patient, the better their chances of death or development of severe health complications are. For instance, the chance of 30-day survival of cardiac arrest drops from 19.5% for 0-6 minutes to 13.8% for 7-9 minutes to 9.4% for 10-15 minutes (Rawshani et al., 2020). Minutes and seconds are of critical importance for these patients lives.

In order to reach emergencies, ambulances are dispatched. Vance County, NC has four ambulances stationed at two EMS stations. In the current situation, there are three ambulances at the center station and one ambulance at the southern station. This decision was made based on the population distribution of the county with the main city of Henderson in the center and more people living in the south than the north, where Kerr Lake is situated. However, it has been observed that the number of calls from the northern region has increased by 37% between 2019 and 2022. This shift in need raises the question of whether or not the county's ambulances could be better allocated and if a northern EMS station should be constructed. Two northern stations have been proposed with one being farther north and the other being closer to the central region. The four alternative arrangements for ambulances we are considering can be found in Table 1.

2 Data and Exploratory Analysis

2.1 Data

Our dataset consists of hundreds of emergency dispatches which were masked and subset from real Vance County ambulance responses. Rows include information on where the call came from, call priority, the ambulance that responded, which station the ambulance came from, which (if any) hospital was the patient transported to, time of response steps (dispatch, starting enroute, arrival at scene, leaving the scene, arrival at hospital, available for next dispatch), and others. The actual dates are not accurate, for privacy. Additionally, we were provided with estimated travel times in seconds from the Google Maps API under four traffic models: unadjusted, pessimistic, best-guess, and optimistic.

We noted that a couple calls were made outside of the borders of Vance County and chose to drop these observations as they are not representative of our goal of optimizing response time to calls within Vance County. A map of all calls can be seen in Figure 1. Additionally, we dropped observations in which no ambulance was reported to have arrived on scene, indicating that no medical attention from the EMS was needed.

2.2 Exploratory Data Analysis

To begin, we wanted to find out how the EMS response times varied by region; specifically, we looked at the response times for calls in the North compared to the rest of the county. Figure 4 displays that calls

from the North do indeed tend to have longer response times, as the median response time is over 2 minutes slower than calls from the Central or South.

We also wanted to determine the distribution of calls and their proximity to each of the four EMS stations (two existing stations and the proposed near north and far north). Figure 5 shows the proportion of calls for which the minimal estimated travel time is that station. From this plot we can see that the Central station is closest to the majority of calls, while the Far and Near North stations are often not the best option. Thus, there is initial evidence towards a Near North station being more advantageous than a Far North station.

Another issue we have to address is the problem of ambulance conflicts. Suppose a call from the south is made but the ambulance at the south station is busy with a prior dispatch. Figure 3 seeks to answer the question of how often these load conflicts occur. For 1.9% of calls, all of the ambulances were busy, and the corresponding calls had to wait for the next available ambulance to be dispatched. For most of the calls (56.1%) all the ambulances were available and ready for dispatch.

Additionally, we wondered if there might be a temporal solution to allocation. For instance, perhaps visitors to Kerr Lake peak on weekends and show a shift in emergency calls towards the north. If this was the case, it would make sense to shift an ambulance up north on weekends. However, in Figure 2 we found that this was not the case. There was no compelling evidence that the proportion of emergency calls from each region changed across time of day or day of the week. We are unable to evaluate seasonal trends from our dataset.

3 Modeling

3.1 Implementation

In order to capture the differences between each of the scenarios, we simulated what would happen with the following methodology. To begin, we removed any duplicate calls in which dispatch times and the ambulance responding were identical as well as ambulances that never arrived at the scene, likely a false alarm. Next, we ordered all calls by dispatch time so we can model availability. Our dispatch rule was that we selected an ambulance from the subset currently available with the lowest estimated travel time. If all ambulances are busy, the next available ambulance is sent as soon as it has been cleared. We repeated this process for all scenarios and all travel time algorithms.

For our plots we use `ggplot2` and for modeling we use the package `lme4`.

3.2 Model Selection

From our simulated scenarios, we sought to model estimated travel times. However, not all scenarios resulted in changes in travel times. These data points caused violations of model assumptions, so we performed a preliminary filtering step that removed any calls in which travel time did not vary across any of the scenarios. After filtering, we had 140 unique calls, each with 5 estimated times (one per scenario).

Next, we created four models, one for each travel time algorithm. Since we are only concerned with how long it takes an ambulance to respond to calls, there is no need to include variables such as distance. We considered adding other predictors, such as time of day or emergency priority, but they did not improve the model, so they were left out. Due to the repetition of rows for the same call under each of the scenarios, we created a variable for call ID upon which we placed a random effect.

Our final model is:

$$\text{EstimatedTT}_{ij} = \beta_0 + \beta_1 \cdot \text{Scenario}_{ij} + u_i + \epsilon_{ij}$$

$$u_i \sim \mathcal{N}(0, \sigma_u^2), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Alternatively, since scenario is categorical, the expanded form is:

$$\text{EstimatedTT}_{ij} = \beta_0 + \beta_1 \cdot S1_{ij} + \beta_2 \cdot S2_{ij} + \beta_3 \cdot S3_{ij} + \beta_4 \cdot S4_{ij} + u_i + \epsilon_{ij}$$

3.3 Evaluation

First, we ensured our simulation’s results were logically consistent by verifying no overlaps in which an ambulance was actively responding to multiple calls at once as well as verifying timing consistency (i.e. dispatch time should be before arrival time).

Next, to evaluate our model, we fit residual plots and looked at other diagnostics. As we can see from Figure 7, modeling assumptions appear to be satisfied and there is no heteroskedasticity. The QQ plot in Figure 7 shows that residuals roughly follow a Normal distribution. We experimented with log transforming our response as well as using a ratio between the baseline and the new scenario but we did not find significant improvements to the model fit.

4 Results

Our model results for the Best Guess traffic algorithm can be found in Tables 2. We estimate that estimated travel time decreases by approximately 45 seconds, on average, in the scenario where we move a Central truck a the Near North station (scenario 3). By contrast, all other scenarios see an increase in expected travel time. Across the four traffic algorithms, the estimated coefficient signals a 45-55 second decrease in estimated travel time for scenario 3. Based on the t-value of Table 2, the estimate is just on the threshold of statistical significance. We also computed how often our dispatch rule allocated an ambulances from different stations in the baseline scenario and scenario 3. We found that scenario 3 tends to send an ambulance from a different station than scenario 0 about 10-11% of the time, depending on the traffic algorithm used.

Additionally, we evaluated how well each scenario deals with load by calculating the optimal station rate, which is the proportion of calls where the ambulance from the closest of the 4 possible stations was dispatched based on estimated travel time. We found that scenario 3 leads to a very small improvement in the optimal station, from 91% in scenario 0 compared to 91.5% in scenario 3 (using Best Guess estimated time). This change is very small, so we conclude that there is no major change in load between the two scenarios. For the remaining scenarios, there is a noticeable decrease in the optimal station rate, meaning they deal with load more poorly than the baseline scenario.

Figure 6 shows the allocation of ambulances by our dispatch rule in each scenario. Conflicts are very rare, but we do see that there are cases where ambulances are dispatched to areas outside of their region. For instance, in scenario 0 in Figure 6, we see a single call from the North that was handled by an ambulance in the South Station. As it would not be ethical to delay aid when an available ambulance from another region could respond, we chose to allow such cross-regional dispatches to proceed.

Similarly, Figure 8 shows the change in estimated travel times (note to self decrease point size) and highlights the plethora of calls for which scenarios do not change travel time estimates. It seems like the increase in travel times in the south under scenarios 1 and 2 did not offset the decrease in travel times in the north. Meanwhile, allocating an ambulance to far north sees too little usage in the north and overall results in travel time increases for central calls.

5 Conclusion, Limitations, and Future Work

5.1 Conclusion

Based on our results, we would recommend Vance County to build the proposed near north station rather than a far north station. The ambulances are best allocated with 1 at south, 2 at central, and 1 at near north. In this arrangement we found a change in estimated travel time of about 45 seconds across travel time estimates compared to the baseline. While this only represents about 10% of calls, any increase in the chance of a patient’s survival is valuable. As such, we determined that a Near North station better serves

the county than a Far North station, and it is better to move a Central truck rather than the South truck. Without information on cost and a focus on lives, we believe a near north station should be constructed.

5.2 Limitations and Future Work

Our analysis makes many assumptions for simplicity. To begin, our simulations assume that all ambulances begin responding to calls from their stations and that there is no time allocated for transport from hospital/caller back to the station. This assumption works well when the utilization of ambulances is low, but in times where utilization is high it's more likely that the starting location of an ambulance would instead be the hospital. In that case, it does not seem to matter which station the ambulance is from as there is only one hospital in Vance County. Next, we are approximating travel time using the Google Maps API which is not representative of traffic patterns, time of day, road closures, weather, and other factors. It's plausible that ambulances with their sirens on are given significant priority the estimates are too high, but it is also plausible that emergencies tend to cause closure and traffic jams so the estimates are too low. Finally, we only have data from a single month and are hence unable to evaluate if scenarios might be better in different seasons. For instance, it is likely that more people will be in the north during summer months to visit Kerr Lake.

As aforementioned in our methodology, we chose a simplistic dispatch rule that greedily selects the ambulance with the shortest travel time. This is because it is unknown how long it will take for "closer" ambulances to finish their prior commitment and waiting when there is an ambulance available would be unethical. Although, there are very few cases in our dataset where all ambulances are all busy. However, calls of a "non-emergency" priority may not be as critical as emergency calls. In these cases, perhaps an alternative dispatch rule would be to ensure an ambulance is always available for emergency calls. In addition to exploring alternate dispatch rules, future work could include an analysis of seasonal trends and other covariates as well as references to actual cost and health outcome data. Perhaps the cost of building a station would be better allocated towards having more ambulances or building another hospital facility. Health outcomes data would also help determine if there is a target transit time as well as prioritization of certain calls over others.

6 Appendix

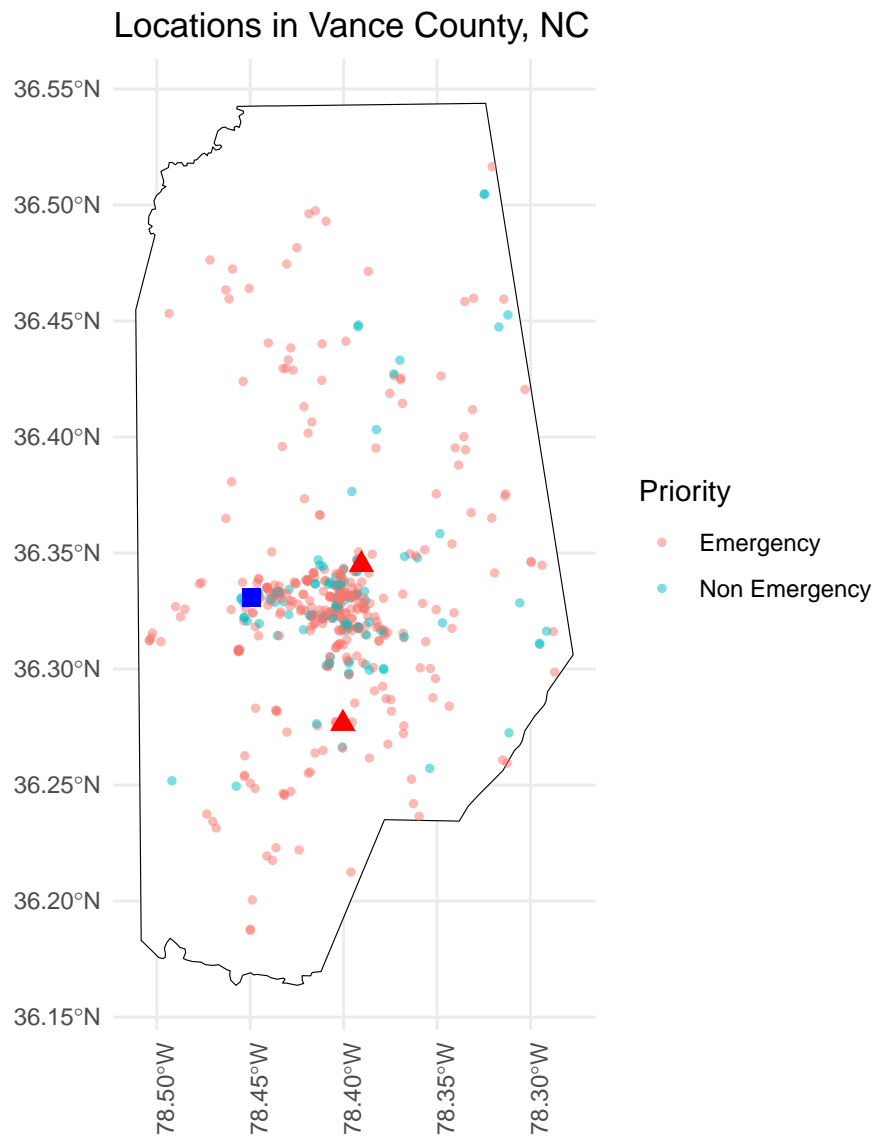


Figure 1: All EMS call locations within Vance County. The red triangles represent the two existing stations while the blue square is the hospital. There does not seem to be any significant relationship between dispatch priority and call location.

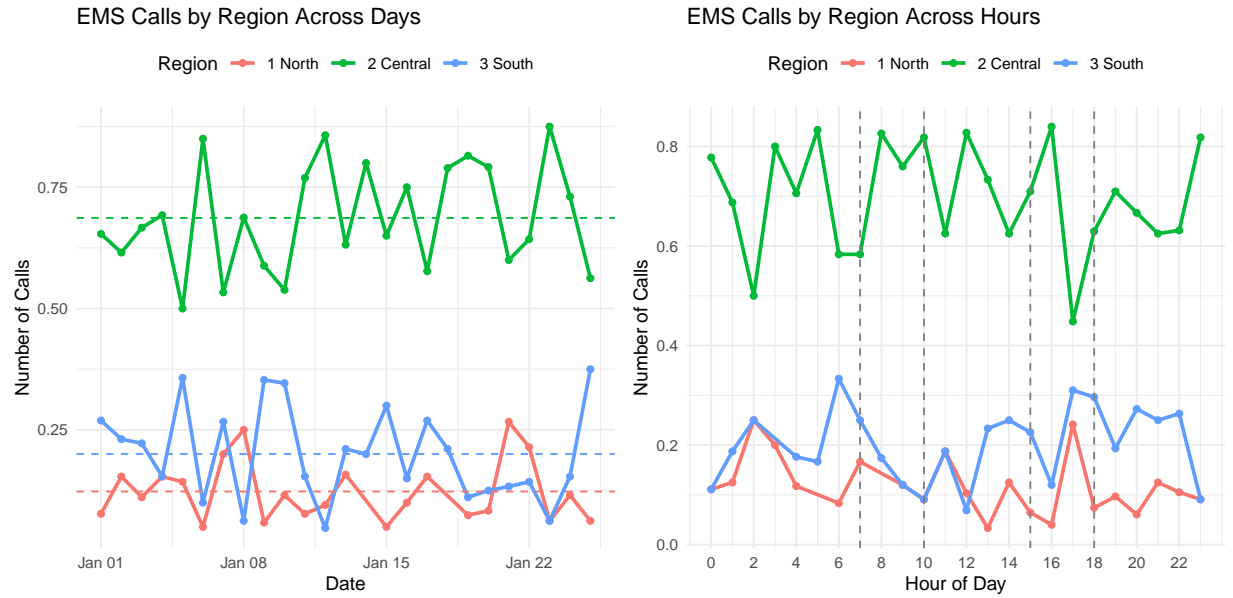


Figure 2: The figure on the left shows calls by date across the three regions. Dotted lines represent the average proportion of calls. The figure on the right shows calls by time of day across the three regions. Dotted lines highlight typical peak traffic hours of 7-10am and 3-6pm. There does not seem to be any clear relationship between date and time.

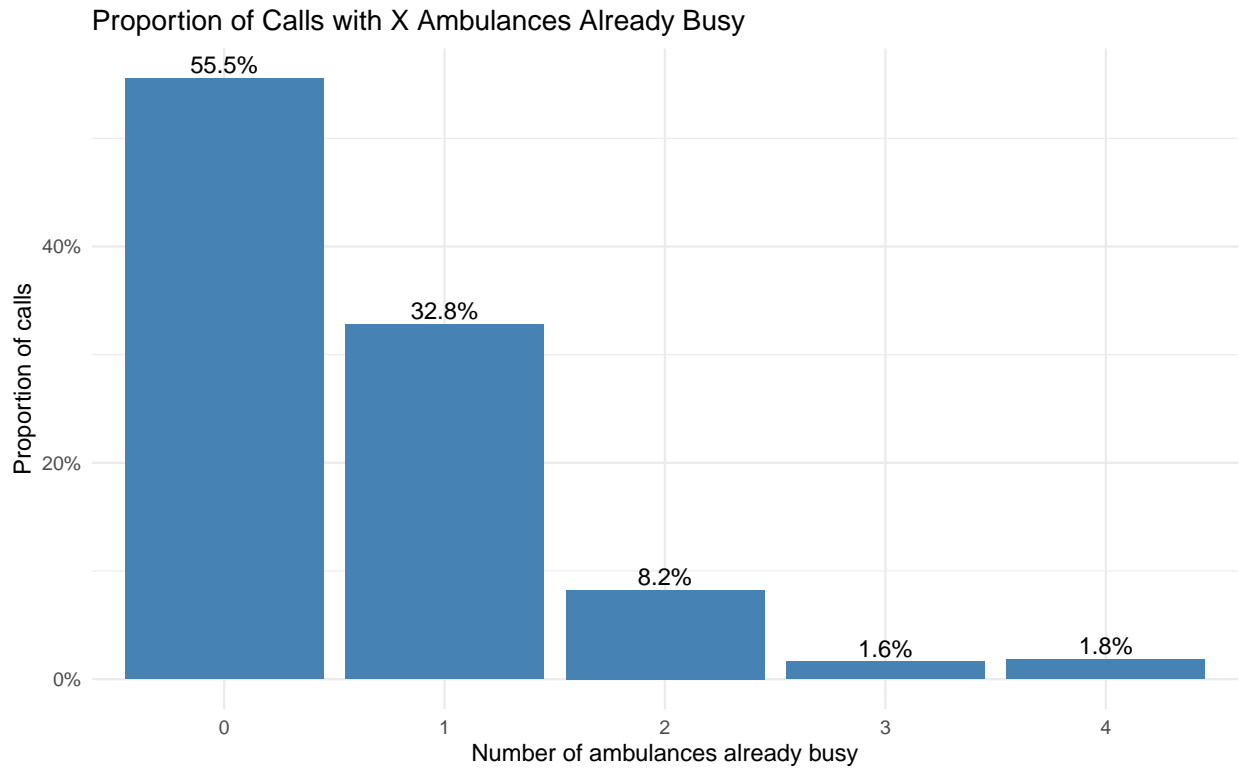


Figure 3: This plot shows the proportion of calls for which different numbers of ambulances are busy. In over half of cases, all ambulances are available. Only less than 2% of time are all ambulances busy.

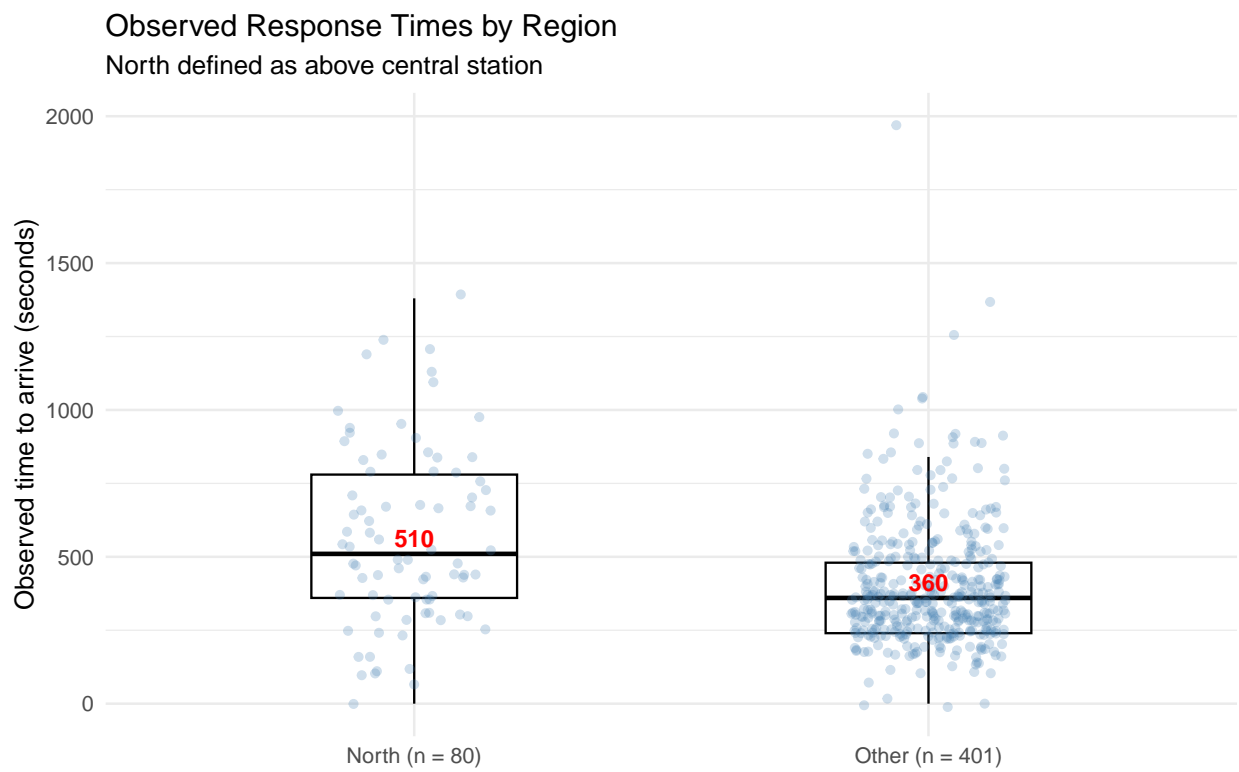


Figure 4: Box plot of response times for northern calls compared to central and southern calls. Northern calls are those for which the latitude is great than that of the central station. Red text highlights the median travel time stratified by call location.

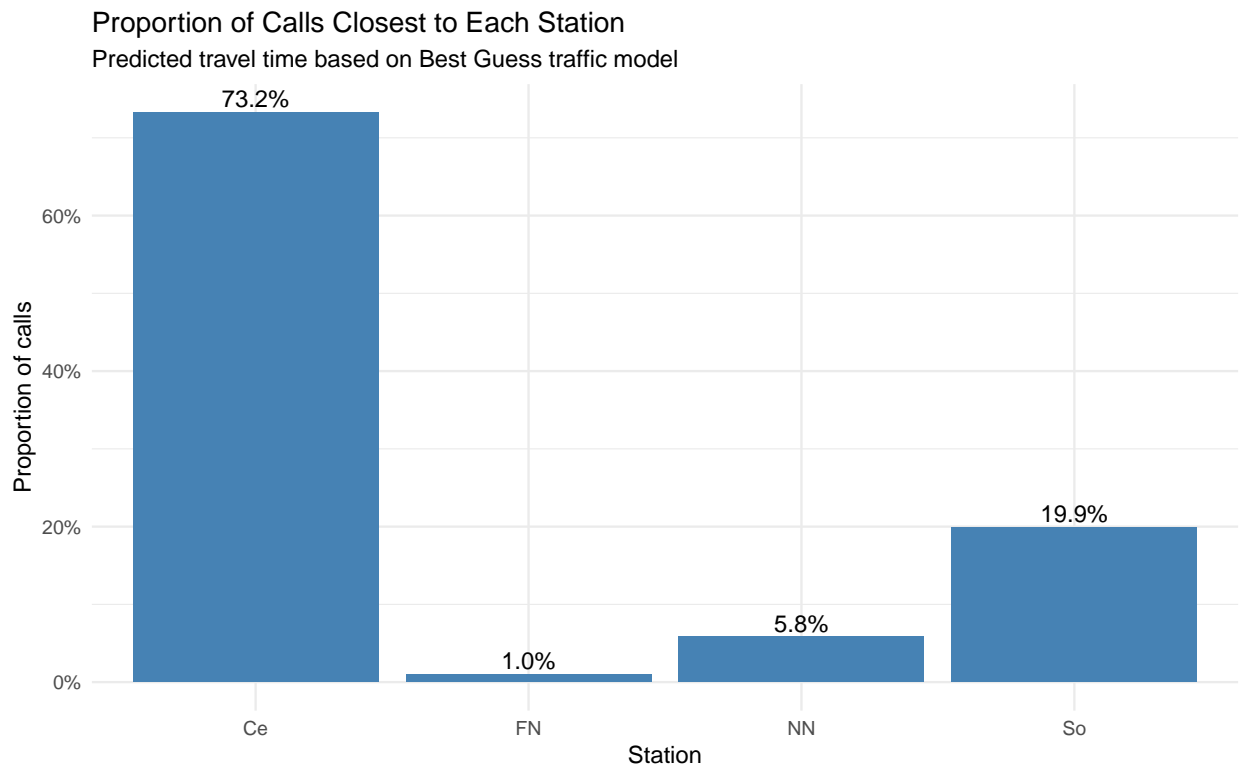


Figure 5: Results of simulation of closest stations, disregarding ambulance availability. In less than 7% of cases the northern stations are the closest. Nearly three quarters of all calls occur in the central region and a fifth of calls occurs in the south.

Dispatched Ambulance by Call Location (Best Guess)

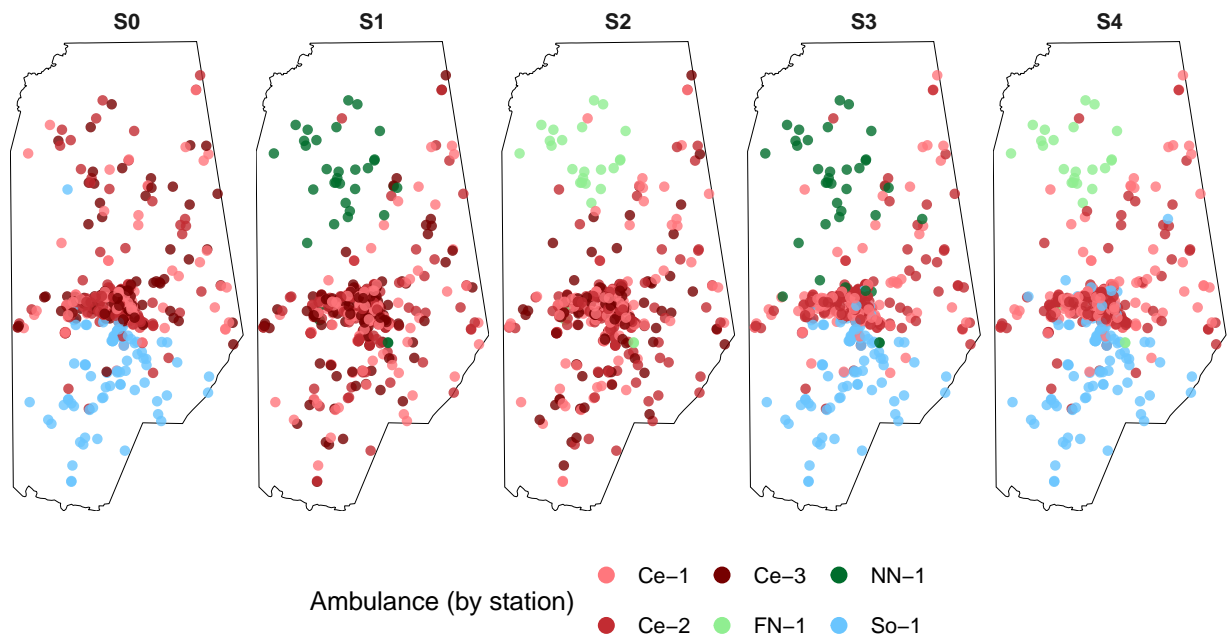


Figure 6: Ambulances dispatched by scenario according to dispatch rule of nearest available ambulance. Notice that many of the central calls experience no change.

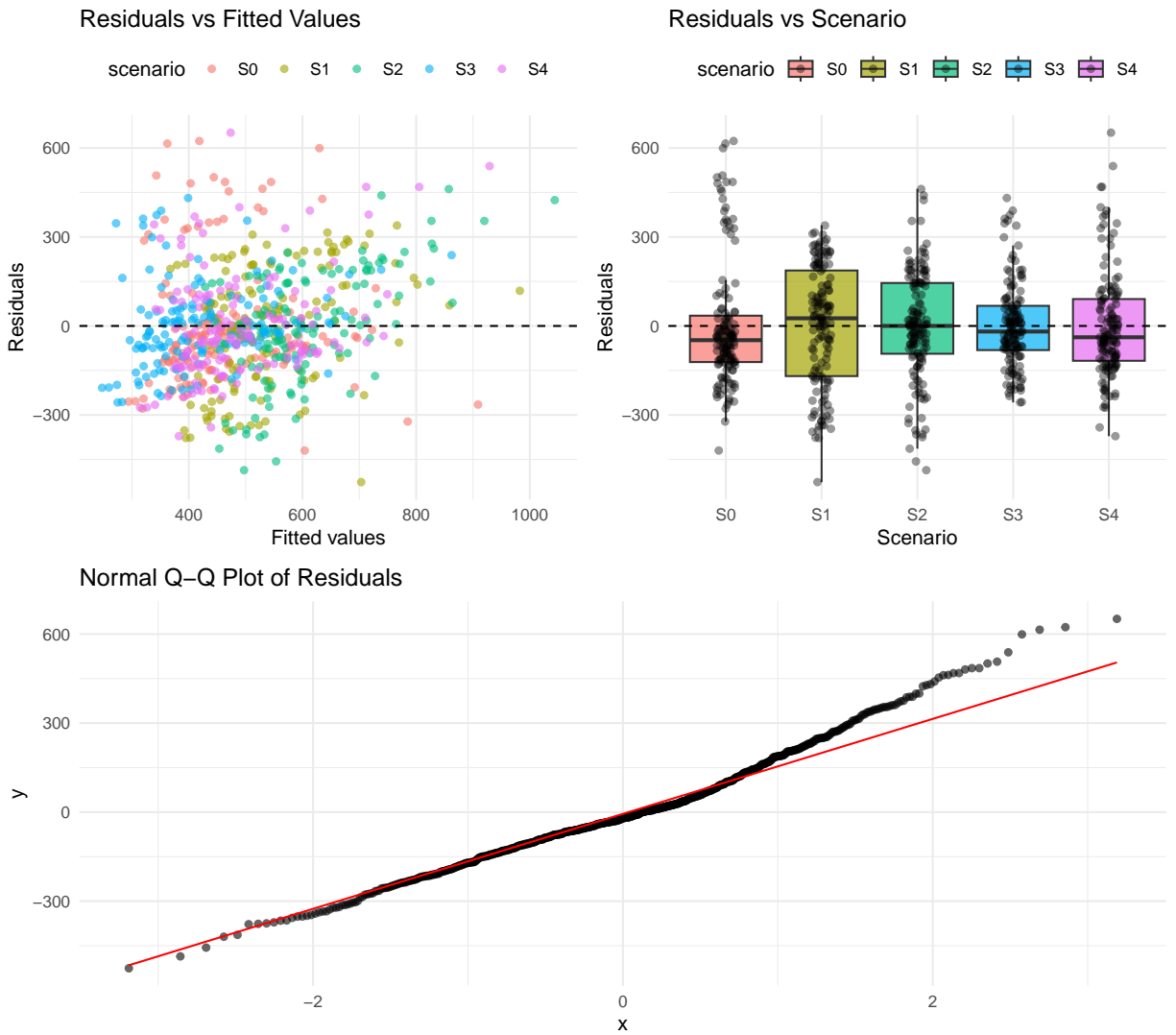


Figure 7: Model Diagnostics

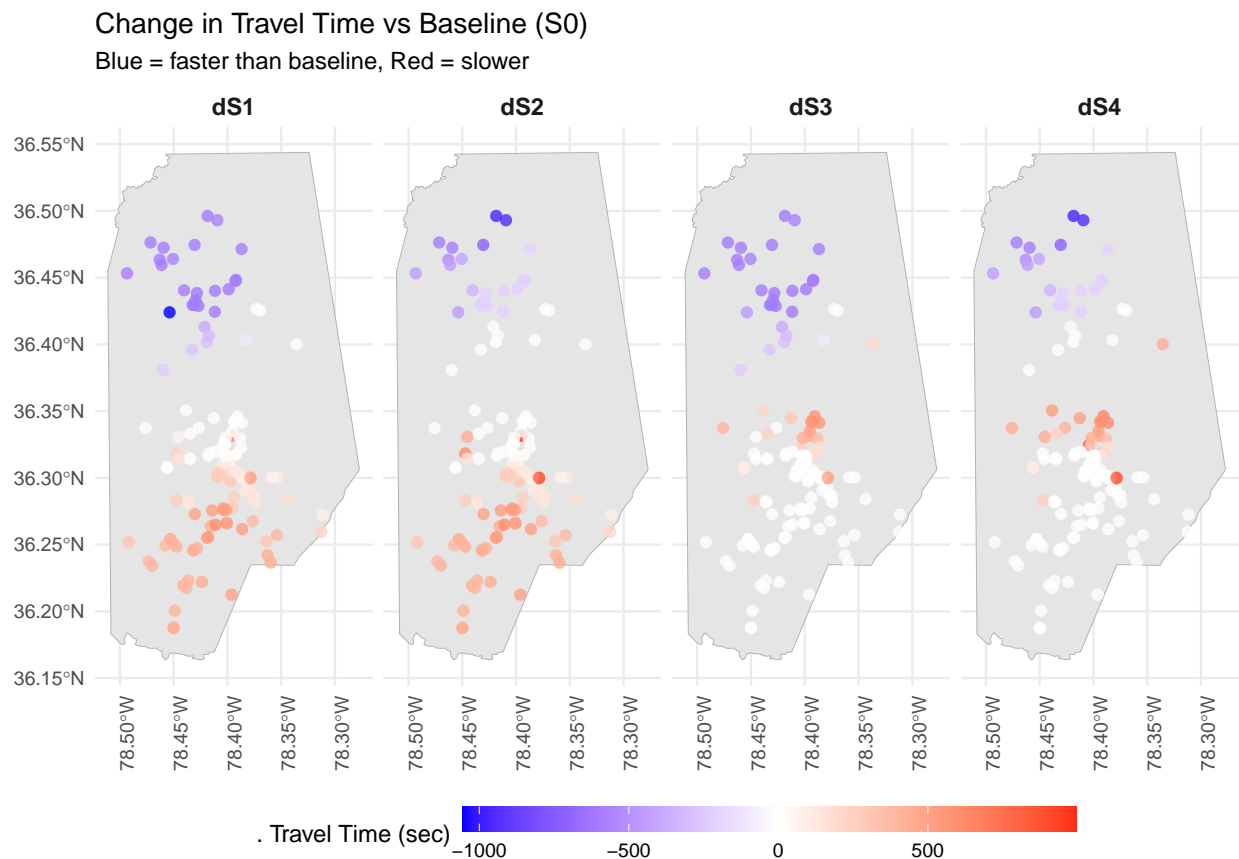


Figure 8: Map of relative changes with respect to the baseline in response times under each of the proposed scenarios. Each point represents a call. Red points represent an increase in estimates, blue points represent a decrease in estimate, and white points represent no change.

Table 1: Ambulance Allocation Scenarios

Scenarios	S0	S1	S2	S3	S4
Far North	0	0	1	0	1
Near North	0	1	0	1	0
Central	3	3	3	2	2
South	1	0	0	1	1

Table 2: Model Results (Best Guess traffic algorithm)

Estimate	Std. Error	t value
467.69	20.06	23.31
73.51	23.82	3.09
134.96	23.82	5.67
-46.61	23.82	-1.96
20.32	23.82	0.85