Supplementary Material for "Metamodel Optimization of a Complex, Rural-Urban Emergency Medical Services System"

Matthew Snyder[a], Byran J. Smucker[a]

[a]*Department of Statistics, Miami University, Oxford, Ohio,*

1. Data Handling Details for Section 3.1

From the raw datasets provided by Chief Lewis, only medical emergencies were included, as this study seeks to improve ambulance response times; Table 1 lists all types of events that were classified as medical emergencies. Calls with missing or illogical assignment, dispatch, enroute, or on scene times were removed, as well as those with missing locations. For the remaining calls, the OpenStreetMap time was computed from the station location to the call location using the `osrm` package in R [1]; only calls within a factor of 1.3 of the OpenStreetMap time were included, per Chief Lewis, to remove those with probable mistakes. This dataset, denoted A, was used whenever reliable times for individual calls were needed.

Emergency Medical Call Codes			
CODE 2	CHEST PAIN	ATTEMPT SUICIDE	ASTINV-LIFT ASSIST
CHOKING	DIFF BREATH	DOMESTIC/MEDICAL	CKPER-PERSON DOWN
LIFEALERT	LIFT ASSIST	OB/GYN EMERGENCY	DIABETIC EMERGENCY
MEDIC	MEDIC ALARM	PER INJ PED	MASLT-MEDIC FROM ASLT
MHEART	MEDIC-MEDIC	PERSON DOWN	MDIAB-DIABETIC EMERGENCY
OVERDOSE	MEDICAL ALARM	PERSONAL INJ	MDIFFB-DIFF BREATH
PSYCH	MHEART-HEART	PI-PERSONAL INJ	MOB-OB/GYN EMERGENCY
SEIZURE	MEDIC FROM ALST	PIP-PER INJ PED	PIHR-PERS INJ HIT/RUN
SHOOTING	MOD-OVERDOSE	RESCUE-RESCUE	PIORV-OFF ROAD VEH PI
SQUAD	MCVA-STROKE	SHOOT-SHOOTING	PIPHR-PER INJ PED HIT/RUN
STABBING	MSEIZ-SEIZURE	STAB-STABBING	SUICA-ATTEMPTED SUICIDE
STROKE	NEED SQUAD		

Table 1: Codes for Medical Emergency Calls

2. Discrete Event Simulation Details for Section 3.2

The distribution of times between call arrivals was based on dataset B, while distribution for the other times were estimated from dataset A. Based on an exploratory analysis, each of these times had heavily right skewed distributions, which led to modeling them with gamma distributions. Plotting the distributions by time of day, weekday, month, season, region, and vehicle type revealed which factors influence the shape of the distributions. Times between call arrivals depended on the time of day and season, while assignment to enroute times depended on time of day and vehicle type. Both vehicle type and the binary city variable were influential for the dispatch to assignment and onscene to clear times. The shape and rate parameters in the gamma density function were estimated for each combination of the influential factors as the maximum likelihood estimators (MLE) using the `fitdistr` function from the `MASS` package. This function uses the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm to numerically optimize the MLEs [2]. For each call, the DES used a randomly generated time from the corresponding distribution.

To calibrate the OpenStreetMaps time to a realistic emergency response time, a linear regression model was fit using dataset A (with the exception that calls responded to by any station, not just the stations of interest, were included). Due to the heavily skewed distributions, the travel times and OpenStreetMap times and distances were log-transformed prior to modeling. Then, along with factors including the time of day, month, season, hour, city, vehicle, and week of year, forward and backward selection procedures were implemented to select the preferred model. In this analysis, both techniques resulted in the same model:

$$\log(travel_i) = \beta_0 + \beta_1 \log(time_i) + \beta_2 \log(dist_i) + \beta_3 tod_i + \beta_4 season_i + \beta_5 vehicle_i + \beta_6 city_i + \\ + \beta_7 \log(time_i) \times log(dist_i) + \epsilon_i$$

where $travel_i$ is the predicted travel time for call i, $time_i$ is the OpenStreetMap time for call i, $dist_i$ is the OpenStreetMap distance time for call i, tod_i is the time of day (early morning, morning, afternoon, evening) for call i, $season_i$ is the season (spring, summer, fall, winter) for call i, $vehicle_i$ is the responding vehicle type (EMS or fire) for call i, $city_i$ is 1 if call i is in the city and 0 otherwise, and $\epsilon_i \sim N(0, \sigma^2)$. (Note that notation has been abused because indicator variables have not been explicitly specified for the categorical variables tod_i, $season_i$, $vehicle_i$, and $city_i$.) Table 2 shows the regression results. The coefficients for both the OpenStreetMap times and distances were positive, indicating longer travel times for longer OpenStreetMap times and distances. However, the interaction between these predictors had a negative coefficient, implying that as the time and distances increase, the travel time does not continue to increase as quickly. Calls located in cities had longer travel times than those in the country, on average, provided that all other predictors are the same. In addition, calls arriving during the morning hours and during the winter were expected to have longer travel times than calls arriving during other times.

Coefficients	Estimate	St. Error	t value	Pr(>| t|)	
(Intercept)	0.053	0.005	10.197	<0.001	***
log(time)	0.919	0.006	151.485	<0.001	***
log(dist)	0.049	0.005	10.648	<0.001	***
todEarly Morning	0.010	0.003	3.637	<0.001	***
todEvening	-0.001	0.003	-0.372	0.710	
todMorning	0.010	0.003	3.343	0.001	***
seasonSpring	-0.003	0.003	-1.108	0.268	
seasonSummer	-0.001	0.003	-0.377	0.706	
seasonWinter	0.013	0.003	4.731	<0.001	***
vehicleFire	-0.025	0.009	-2.881	0.004	**
city	0.035	0.003	12.528	<0.001	***
log(time)× log(dist)	-0.007	0.001	-6.600	<0.001	***
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	
Residual standard error: 0.140 on 19163 degrees of freedom					
Multiple R-squared: 0.963, Adjusted R-squared: 0.963					
F-statistics: 4.606e+04 on 11 and 19613 DF, p-value: <0.001					

Table 2: Multiple linear regression summary output, rounded to three decimal places.

Using this regression model, the predicted log travel time and standard error of its 95% prediction interval between each call and all stations in the configuration were computed. Then the values corresponding to the nearest station were used as the mean and standard deviation of a normal distribution, from which a number was randomly selected and exponentiated to obtain the travel time used in the DES.

This simulator took just under 1 minute to run one iteration for the one-station setting, just under 2 minutes for the two-station setting, approximately 8 minutes for the five-station setting, and over 20 minutes for the twelve-station setting. These times were based on one core of a 3.4gHz Intel Core i5 processor with 8 gigabytes of RAM, and scaled linearly for each additional iteration.

3. DES Validation for Section 3.2

The discrete event simulation used for this analysis is a model of the actual EMS system in St. Louis County, Minnesota. To evaluate the accuracy of the created DES in simulating the actual system, an adjusted DES was created for the one-station, two-station, five-station, and twelve-station settings that uses a different sample of call locations.

3.1. Data and Methods

The difference between the simulators described in Section 3.2 of the article and the simulators used in this validation was the set of locations where calls originated. Rather than sampling from all locations in dataset C, only those locations found in dataset A were used in order to create a fair comparison to the true response times in dataset A. A full comparison using all locations from dataset C could not be conducted due to the missing and inaccurate times for many of the calls. Then the simulator was run 100 times using the current station configuration, and both the overall density and the 90th percentile of response times were compared to the actual response times found in dataset A.

3.2. Results

3.2.1. One Station

For the one-station setting, the mean 90th percentile of response times from the 100 simulated runs was 11.73 minutes with a 95% confidence interval of (11.69, 11.76). The actual 90th percentile response time from the 6,240 calls in dataset A was 11.72 minutes. Figure 1 shows the distribution of response times from the actual data in red, and from one run of the simulation in blue. Both distributions are heavily right skewed with peaks just under five minutes.

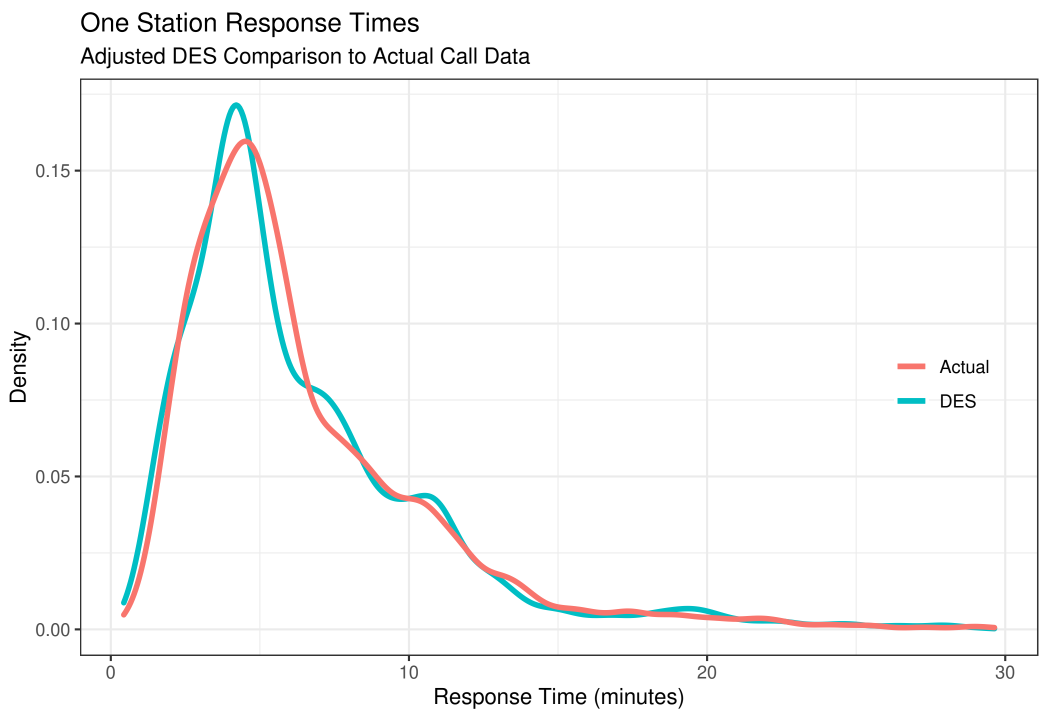


Figure 1: Example density plots of simulated and actual response times for the one-station setting.

3.2.2. Two Station

The mean 90th percentile of response times from the 100 simulated runs of the two-station version was 11.32 minutes with a 95% confidence interval of (11.29, 11.34). Based on the 7,129 calls in dataset A, the actual 90th percentile response time was 12.01 minutes, approximately 30 seconds higher. Figure 2 shows the distribution of response times from the actual data in red, and from one run of the simulation in blue. Although there is some difference in the 90th percentiles, the distributions are again right skewed and match fairly closely.

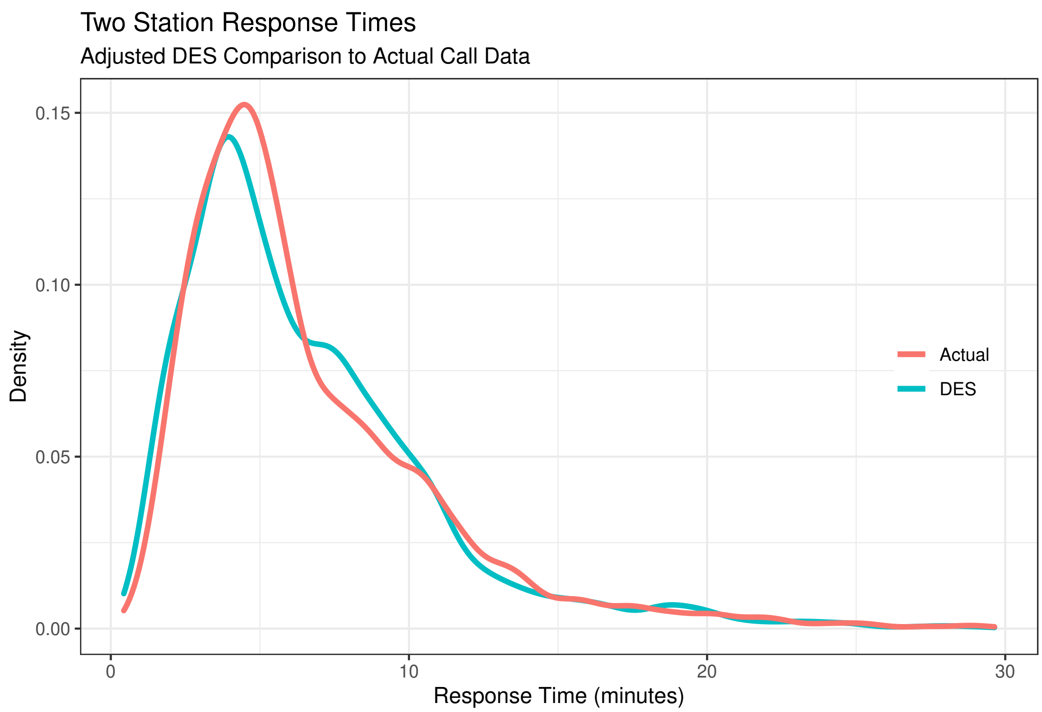


Figure 2: For the two-station setting, example density plots of simulated and actual response times.

Figure 3 shows the spatial call distribution, color coded by the responding station, for a simulated run on the left and the actual data on the right. Since the simulation was designed to assign a call to the closest station, provided that it had an available ambulance, there is a clear distinction between calls responded to by the Virginia and Eveleth stations. In the real data, the distinction is less clear, with Virginia responding to many calls in the southwest and southeast corners of the region even though the Eveleth station is closer. An intentional choice to use distances rather than jurisdictions in the DES is likely the cause of differences in the 90th percentile, since most of the differences in the responding station occur in the outlying areas.

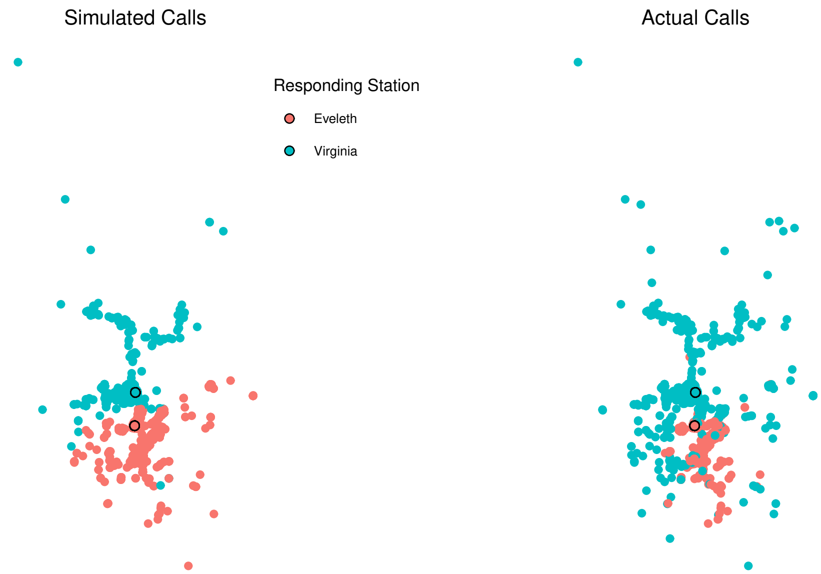


Figure 3: Comparison of simulated and actual call locations and responding stations from the two-station setting.

3.2.3. Five Station

For the five-station setting, the mean 90th percentile response time of the 100 simulated runs was 11.13 minutes with a 95% confidence interval of (11.11, 11.15). The actual 90th percentile of response times for these five stations was 11.15 minutes, contained within the confidence interval. Figure 4 shows similar right-skewed distributions for the overall response times, although the simulated curve is shifted slightly to the left, indicating that there are more simulated responses under five minutes than actual responses.

Figure 5 shows which call locations were responded to by which stations for a simulated run on the left and the actual data on the right. In general, the DES matches the real data, with the exception of the Virginia and Eveleth discrepancy discussed in the two-station setting. However, there are more calls and more possible outlying regions for the five-station setting, so this discrepancy is less impactful on the 90th percentile, resulting in similar metrics.

3.2.4. Twelve Station

The adjusted DES for the twelve-station setting has similar results as the five-station setting. The mean 90th percentile of response times over 100 simulated runs was 12.91 minutes with a 95% confidence interval of (12.57, 13.29). The actual 90th percentile fell within this range, at a value of 12.99 minutes. Figure 6 shows the distributions of response time between a simulated run and the actual times, showing similar patterns as the five-station setting.

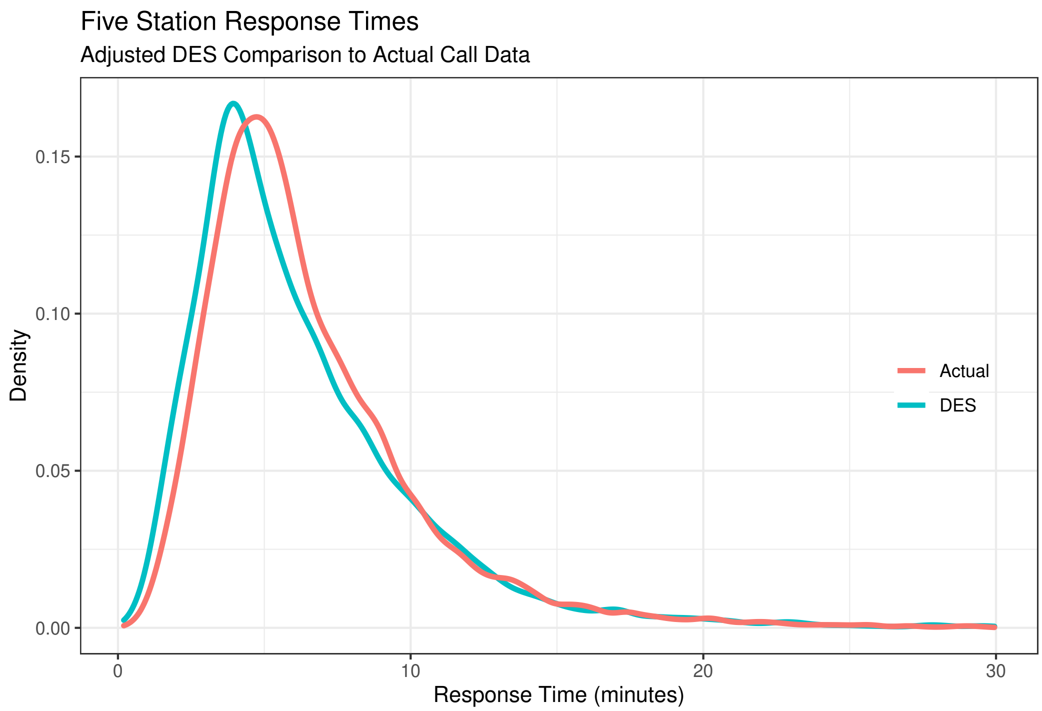


Figure 4: For the five-station setting, example density plots of simulated and actual response times.

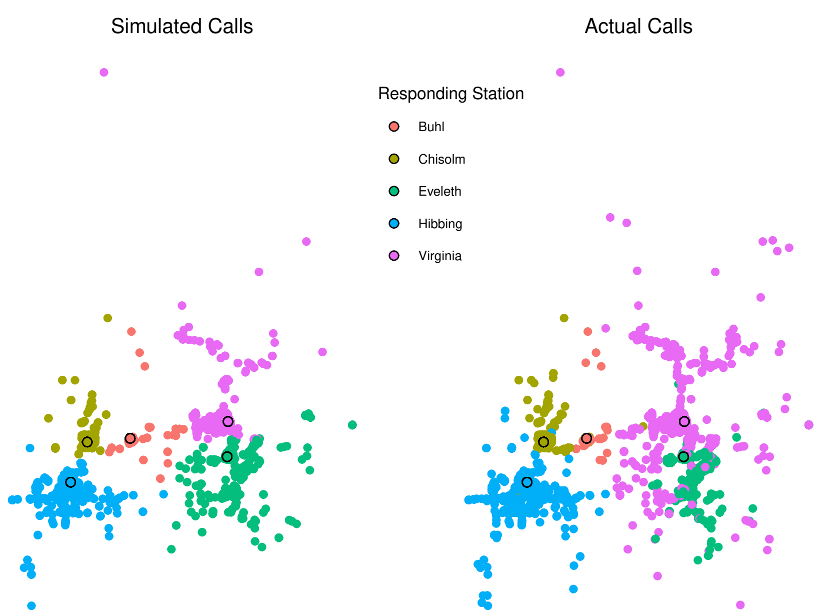


Figure 5: Comparison of simulated and actual call locations and responding stations from the five-station setting.

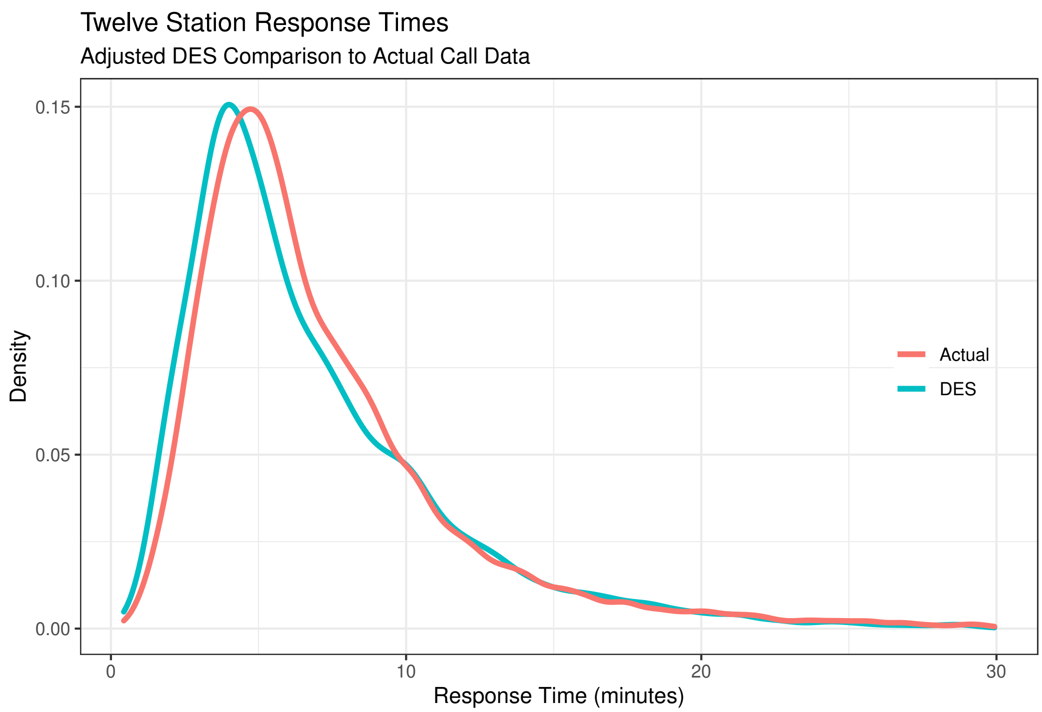


Figure 6: Example density plots of simulated and actual response times for the twelve-station setting.

Finally, Figure 7 shows the spatial location of calls and their responding stations, with the most notable difference being the discrepancy between the Virginia and Eveleth stations described in the two-station setting.

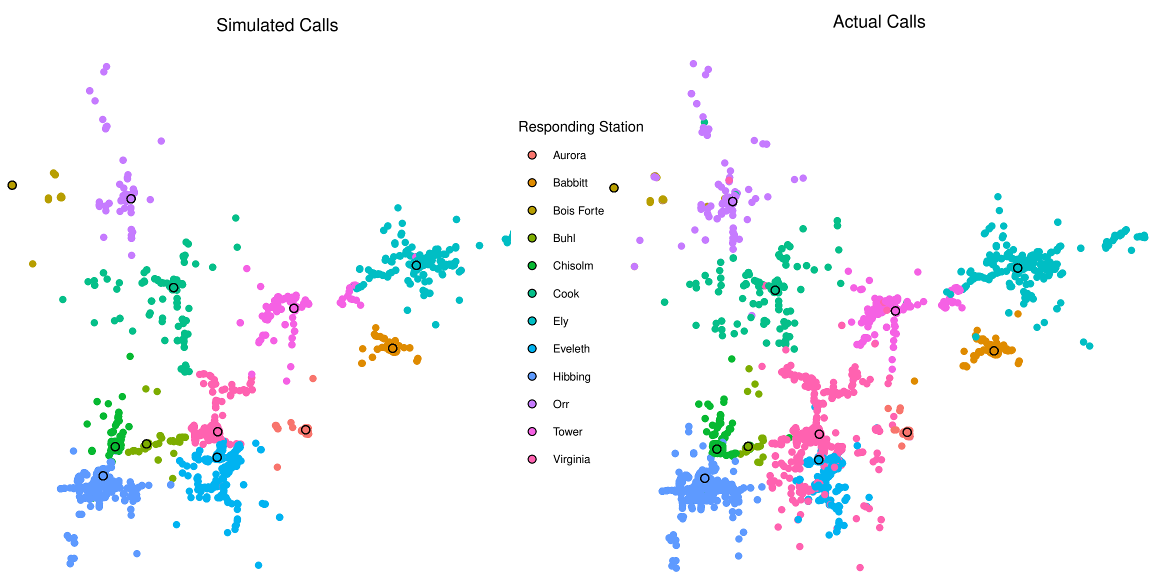


Figure 7: Comparison of simulated and actual call locations and responding stations for the twelve-station setting.

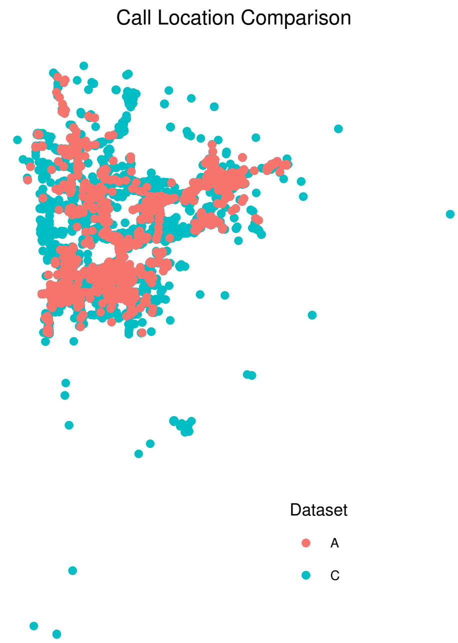


Figure 8: Comparison of call locations from Datasets A and C.

3.3. Discussion

Overall, the adjusted DES appears to model the actual call data fairly well, especially with respect to the 90th percentile of the response times. The systematic differences for the two-station setting, likely due to the distance-based method of assigning stations, were alleviated in the more complex systems. However, it is important to recall that this DES only uses a subset of the call locations found in the full dataset. Figure 8 below shows the call locations in dataset A for the twelve-station setting in red, used for the DES in Section 3.2.4, and the call locations in dataset C in blue, which were used for the DES in the main analysis.

Thus, using all locations in dataset C for the full analysis is an extrapolation to this validation study. Even then, these results provide some evidence that the DES accurately models the true EMS system, and all locations from dataset C were included in the full analysis with the hopes of obtaining a more realistic simulation.

4. Sampling of Station Configurations for Section 3.3

The `density.ppp` function of the `spatstat` R package was used to construct a spatial density for calls based on dataset C. For every point x in a 1024-by-1024-pixel region, the intensity $\lambda(x)$ was calculated based on all points x_i in the neighborhood of x as $\lambda(x) = \frac{\sum_{i=1}^{n} \delta(x-x_i)}{p(x)}$, where $\delta(x)$ is the probability density function of a normal distribution with a mean of zero and standard deviation of σ. The smoothing bandwidth for the kernel estimation was determined using Scott's rule of thumb so that σ is proportional to $n^{-1/(d+4)}$, where n is the number of points and d is the number of spatial dimensions; this bandwidth method is beneficial when estimating gradual trends in the region [3], which is the goal of this analysis. The end correction was computed as $p(x) = \int_R \delta(x-u)du$ where R is the design region. Finally, the density at point x was calculated as $\rho(x) = \frac{\lambda(x)}{\sum \lambda(x_i)}$ and these densities were used for the weighted sampling.

5. Sampling Approach Exploration for Section 3.3

5.1. Introduction

In order to train the metamodels for this analysis, a sample of station configurations must be generated and run through the DES. Since each DES run takes computation time, the goal is to select a sampling method that minimizes the amount of data needed while still maintaining accurate and informative results. Intuitively, a well-placed configuration has stations near the majority of the calls but is also spread throughout the region so that it can reach all locations relatively quickly. Thus, four sampling techniques were explored that balance various weighting and distance-based approaches in this brief, exploratory study.

5.2. Data and Methods

The five-station setting with 2,000 samples was selected for this comparison in order to account for the complexity of multiple stations but not require extensive DES run times. Four sampling techniques – Unweighted, Weighted, Constrained, and Adjusted Bandwidth – were compared.

For the Unweighted sampling approach, a convex hull of the call locations found in dataset C was constructed using the `chull` function of the `grDevices` package in base R [4]. Then a sample of 2,000 configurations was randomly generated, each consisting of five points, using the `spsample` function from the `sp` package [5; 6], resulting in the unweighted sample.

Next the Weighted samples were generated as described in Section 4 using the density of call locations based on Scott's rule of thumb for the bandwidth, but extended for the five-station setting by sampling five locations for each configuration. The Constrained sampling approach is described in Section 3.3 of the article where a four mile constraint on the locations within a configuration is incorporated. Finally, the Adjusted Bandwidth method was implemented in the same way as the Weighted method, but multiplying the bandwidth, based on Scott's rule of thumb, by ten. The larger bandwidth resulted in a more even distribution rather than strong peaks in a few areas, encouraging the configurations to be spread out more broadly across the region.

The average and average minimum distances between stations were compared across the 2,000 configurations for the four approaches. Then, metamodels were fit and station locations optimized for each approach, and the proposed configurations were validated on the DES, as described in Section 3 of the article. In addition to considering the model fit, the best approaches were chosen based on the improvement in 90th percentiles for the proposed configurations compared to the current.

5.3. Results

Table 3 displays the average and average minimum distances between stations across the 2,000 configurations for the four methods. On average, the Unweighted method resulted in the most spread out configurations, while the Weighted method had much tighter configurations. The Constrained technique increased the average minimum distance between stations from just over two miles (3.3 km) to just over five miles (8.2 km) compared to the Weighted method. The Adjusted Bandwidth method increased these distances even more, but not nearly as much as the Unweighted method.

Method	**Average Distance between Stations (miles)**	**Average Minimum Distance between Stations (miles)**
Unweighted	62.7	18.4
Weighted	12.2	2.0
Constrained	14.9	5.1
Adjusted Bandwidth	19.9	6.2

Table 3: Average Distances between Stations by Sampling Method

After running the DES for each of the 2,000 configurations and each of the four methods, a metamodel was fit for each method. Figure 9 shows the average R^2 values from the cross-validation procedure, which show that the Unweighted and Weighted methods had values between 65% and 75%, while the Constrained and Adjusted Bandwidth methods had much lower values, around 35% and 50%, respectively.

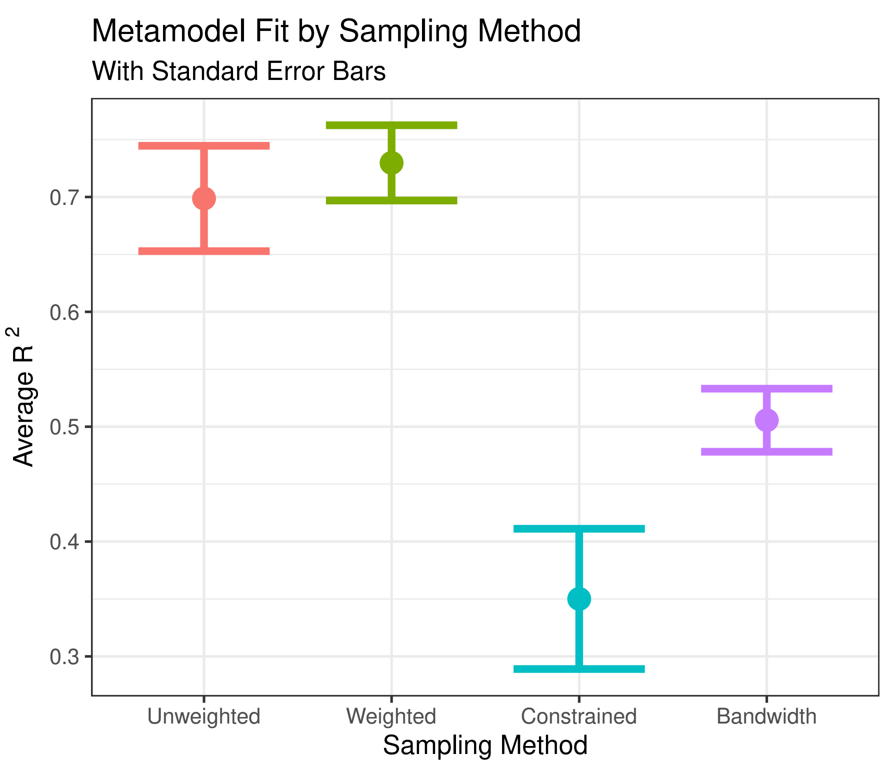


Figure 9: Metamodel fit by sampling method using the five-station setting.

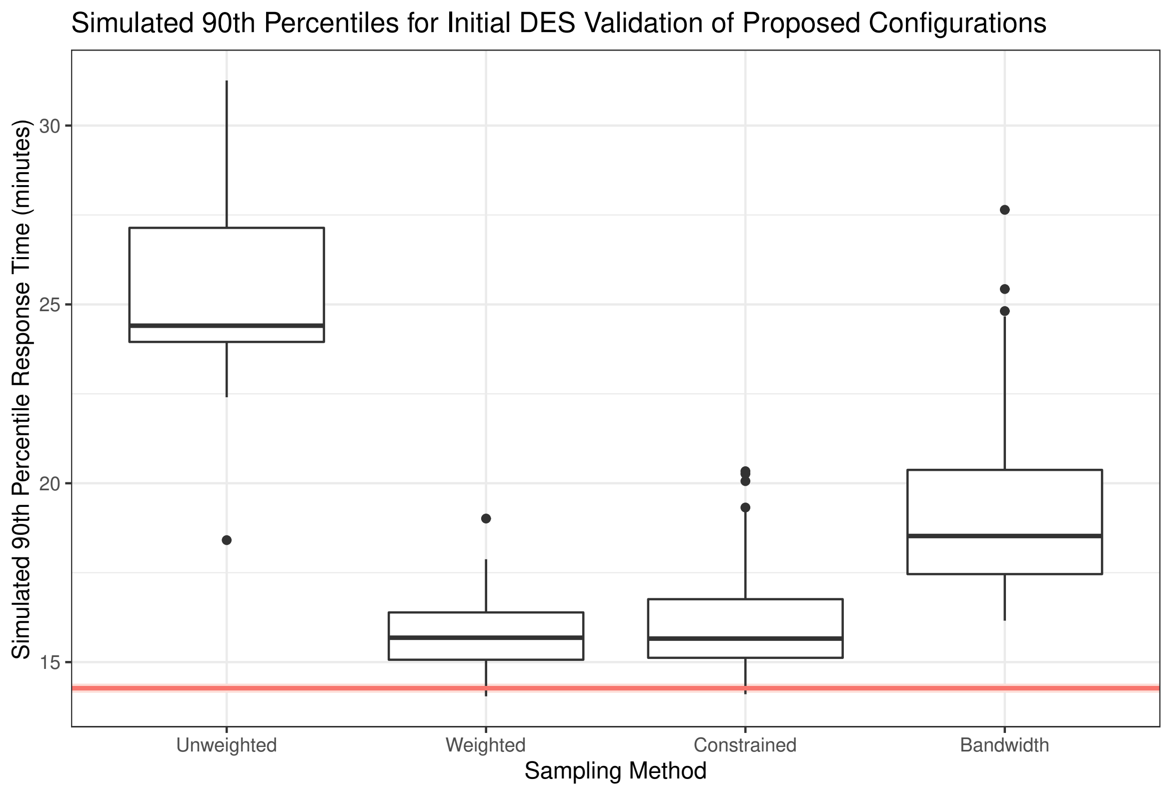


Figure 10: Boxplots of inital DES validation of proposed configurations by sampling method for the five-station setting. The red line represents the current configuration's average simulated 90th percentile along with 95% confidence bands.

Next, station configurations were optimized using each metamodel with 100 different initial configurations and the resulting optimal stations were validated one time on the DES. Figure 10 shows boxplots of the simulated 90th percentile response time for the proposed configurations for each method, with a solid line representing the mean 90th percentile response time of the current configuration. This shows that all proposed configurations using the Unweighted and Adjusted Bandwidth methods had simulated 90th percentiles well above the current configuration. The Weighted and Constrained methods resulted in configurations with similar simulated 90th percentile distributions and were much closer to the current configuration's simulated 90th percentile. However, only one configuration for the Weighted case and four configurations for the Constrained case had simulated 90th percentiles below the upper 95% confidence interval for the 90th percentile of the current configuration.

5.4. Discussion

The Unweighted and Adjusted Bandwidth methods resulted in configurations spread throughout the region, while the Weighted and Constrained methods encouraged stations to be located in areas with high call densities. While the metamodels built on the Unweighted and Weighted samples had much better fit than the other methods, the Weighted and Constrained methods resulted in much better optimal configurations, in general. So, even though a sample of size 2,000 is insufficient to consistently improve upon the current configuration, this exploratory study demonstrates that weighting the configurations based on the call density is beneficial for finding configurations with improved 90th percentiles. Thus, the Constrained method was implemented for the case studies with more than two stations in the full analysis, but a more rigorous comparison should be performed to obtain more reliable results.

6. Random Forest Details for Section 3.3

Random forest models are composed of several regression decision trees. A regression decision tree starts by splitting the predictor space into j non-overlapping regions R_i, $i = 1, 2, \ldots, j$, which are decided such that the residual sum of squares, $\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2$, is minimized. Here the predictors are the latitudes and longitudes of each station in the configuration, y_i is the 90th percentile from the simulation, and $\hat{y}_{R_j}$ is the fitted 90th percentile value from the metamodel. This greedy algorithm then repeats, continuing to split into smaller regions to minimize the sum of squares.

Random forest models fit multiple decision trees and average the predicted value from each tree in order to reduce the variance of the prediction. However, at each step, only a sample of m predictors are considered for splitting, in order to prevent all trees from following the same path [7]. For this analysis, 10-fold cross-validation was used to select the value of this tuning parameter that maximizes the R^2 of the model. This value is computed as the squared correlation between the fitted and observed values, so that $R^2 = \left(\frac{n \sum (y_i \hat{y}_i) - \sum y_i \sum \hat{y}_i}{\sqrt{[n \sum y_i^2 - (\sum y_i)^2][n \sum \hat{y}_i^2 - (\sum \hat{y}_i)^2]}} \right)^2$, where n is the number of data points, $\hat{y}_i$ is the fitted value from the metamodel, and y_i is the observed 90th percentile from the simulation [8].

For the one- and two-station settings, models were fit and compared with $m = 1, \ldots, p$, where p is the number of predictors. For the more complex five- and twelve-station versions, values of m ranged from 1 to $p/3$, as recommended by Liaw and Wiener (2002), to simplify the computational costs [9].

7. Particle Swarm Optimization Details for Section 3.4

As introduced in Section 3.4 of the article, the 2007 version of standard particle swarm optimization (SPSO 2007), as defined by Clerc [10], was implemented. The first step is the initialization of the swarm, consisting of $S = 10 + \lfloor 2\sqrt{d} \rfloor$ particles where $d = 2K$, $K = 1, 2, 5, 12$, is the number of dimensions. Each particle, $i = 0, 1, 2, \ldots, S - 1$, is assigned a random starting configuration within the search space and a random initial velocity. An adaptive random topology is used for the neighborhood of each particle, so that each particle informs three random particles in the swarm.

Once the initialization is complete, the iterative process begins, starting at time step $t = 1$. Each particle moves around the search space based on a combination of its previous velocity, its previous best configuration (in terms of minimizing the objective function), and the best configuration previously found within the particle's neighborhood. At the new particle location, the metamodel is evaluated and the current configuration $x_i(t)$, best configuration $p_i(t)$, and best configuration in the neighborhood $n_i(t)$ are updated.

Note that $x_i(t)$, $p_i(t)$, and $n_i(t)$ each have d components, one for each latitude and longitude of every station in the configuration. Using these values, the velocity $v_i(t)$ is then updated coordinate-by-coordinate as $v_{i,d}(t+1) = \frac{1}{2\log(2)} v_{i,d}(t) + c_1 \times (p_{i,d}(t) - x_{i,d}(t)) + c_2 \times (n_{i,d}(t) - x_{i,d}(t))$ where c_1 and c_2 are random numbers taken from a Uniform$(0, 1/2 + \log(2))$ distribution. The exploitation constant ($\frac{1}{2\log(2)}$) and local and global exploration constants $(1/2+\log(2))$ are standard for SPSO 2007. If the velocity forces the particle outside of the search space, the particle is moved to the boundary and the velocity is set to zero.

This definition of the velocity, which incorporates both past information and random components, allows the procedure to converge to the optimum while still exploring the search space. Further, for each time step, the particles move in a random permutation in order to prevent premature convergence. This process stops after 50 iterations pass without any improvement in the objective function.

8. Details for One Station Results (Section 4.1)

This PSO algorithm was used to optimize the station locations using each of the fitted metamodels from Section 4.1 of the article 100 times using different initializations. See an example of the convergence rate in Figure 11. There are significant improvements in the first 30 iterations, after which the change is minimal.

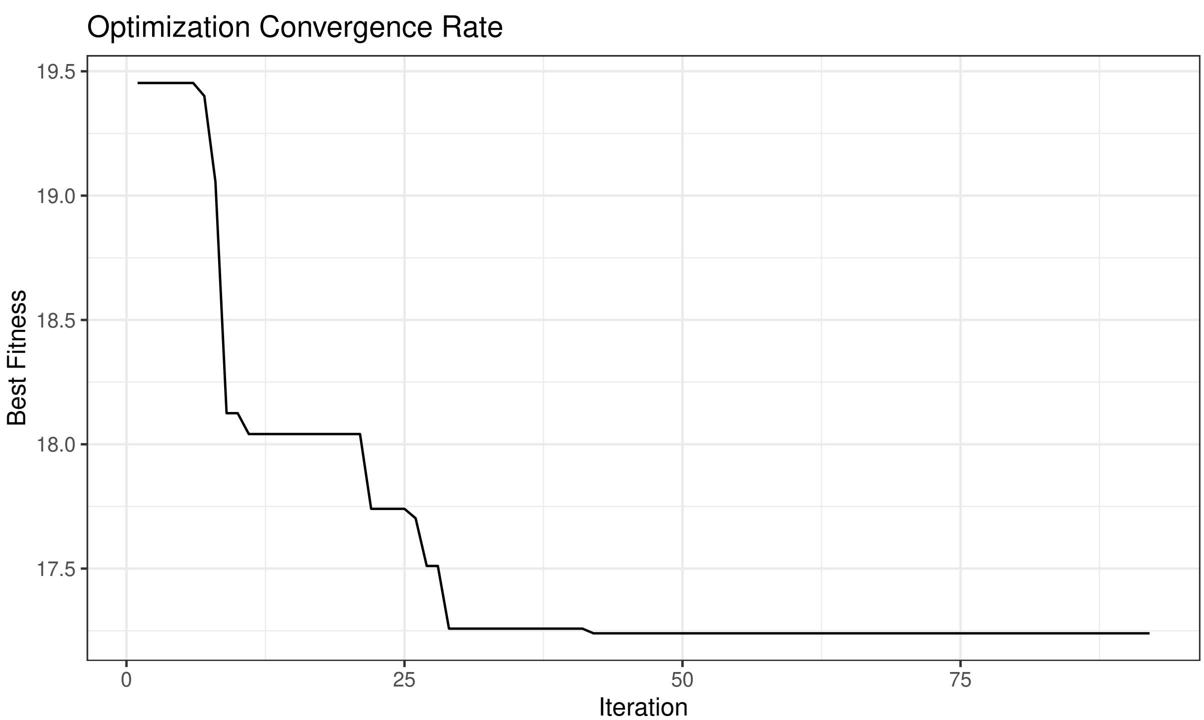


Figure 11: An example of the convergence of the PSO algorithm, for the one-station setting in which the metamodel was built based upon a sample size of 1,000.

Figure 12 shows the resulting station locations from all 100 optimization procedures for each metamodel, along with the current station location in red and all call locations in black. For a given sample size, most

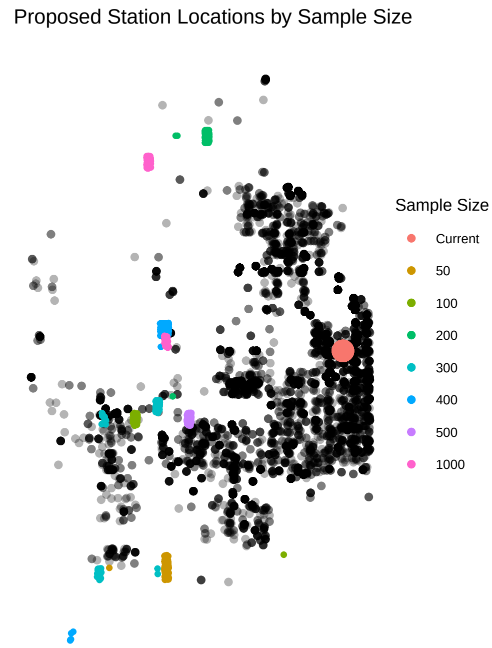


Figure 12: For the one-station setting, locations identified by the various PSO runs, by sample size.

locations are in similar regions, and all are west of the current location. An evaluation of these proposed configurations is presented in Section 4.1 of the article.

9. Details for Two Station Results (Section 4.2)

Gamma distributions based on the same factor levels described in Section 3.2 of the article were used to represent the times between call arrivals, and the times from dispatch to assignment, assignment to enroute, and arrival to clear for each call. Table 4 shows the parameter values for each combination of factor levels in this set of data. In addition, the same regression model was used for the travel time that is described in Section 2 for this two-station setting. Using these distributions, the DES was constructed and a validation is presented in Section 3 of this Supplementary Material.

In order to fit a random forest metamodel over this DES, weighted samples of six incremental sizes were generated as inputs, as described in Section 3.3 of the article. Figure 13 shows four example configurations with stations in blue and call locations in black; due to the high call density in the center of this region, many stations were placed close together near the current Virginia station. Then, for each configuration in each sample size, the DES was run for one year and the 90th percentile response time was computed.

Next the particle swarm optimization procedure was performed using 100 different starting configurations for each of these six metamodels. Figure 14 shows an example of one procedure's convergence for the metamodel built on a sample of 5,000 configurations. There are significant improvements in the first 30 iterations, additional decreases until the 75th iteration, and minimal improvement after. Although similar

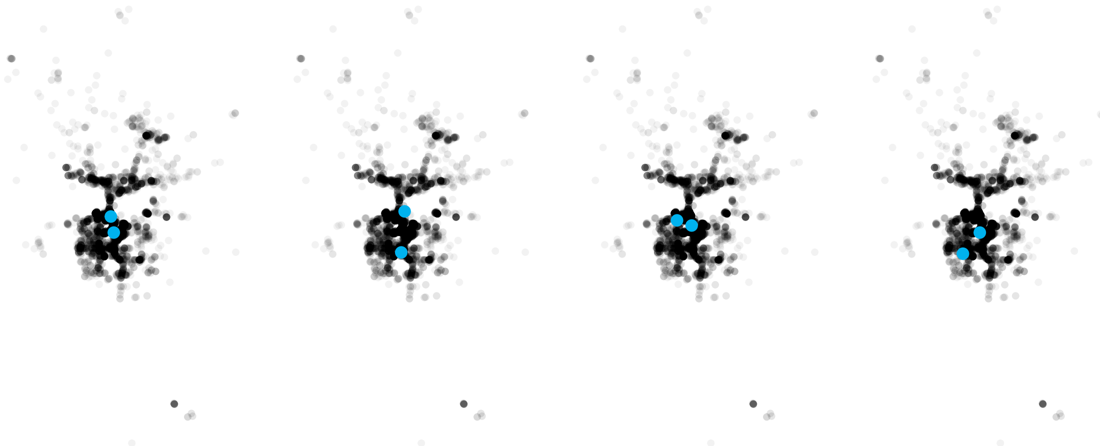


Figure 13: For the two-station setting, example configurations used as inputs for the metamodels.

Time Difference	**Factor Levels**	**Early Morning**		**Morning**		**Afternoon**		**Evening**	
		Shape	**Rate**	**Shape**	**Rate**	**Shape**	**Rate**	**Shape**	**Rate**
Call Arrival	**Spring**	1.069	0.007	0.784	0.006	0.897	0.012	1.101	0.012
	Summer	1.322	0.008	0.824	0.006	1.013	0.013	1.093	0.013
	Fall	1.238	0.008	0.766	0.006	0.969	0.124	1.172	0.014
	Winter	1.232	0.008	0.753	0.007	1.017	0.015	1.156	0.014
Assignment to Enroute	**EMS Vehicle**	1.466	0.731	0.817	0.440	0.866	0.686	1.030	0.719

		In City		**Not in City**	
		Shape	**Rate**	**Shape**	**Rate**
Dispatch to Enroute	**EMS Vehicle**	2.511	1.085	1.844	0.612
Onscene to Clear	**EMS Vehicle**	1.841	0.032	1.632	0.023

Table 4: For the two-station setting, parameter values for the gamma distributions used in the DES.

to the one-station setting, this optimization procedure required more iterations to reach the termination criterion.

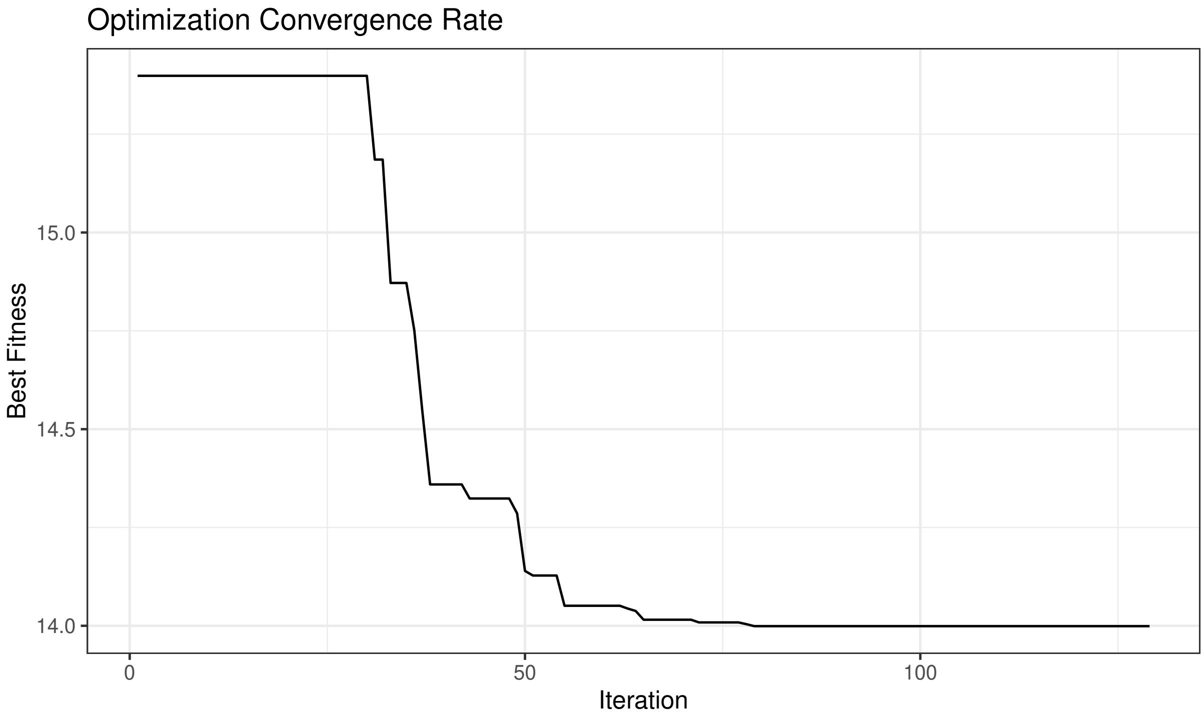


Figure 14: An example of the convergence of the PSO algorithm, for the two-station setting in which the metamodel was built based upon a sample size of 5,000.

The proposed configurations were then validated on the DES, with results displayed in Figure 7b of the article. Note that in this figure, 95% confidence intervals are not shown due to the large number of different, proposed configurations. The size of each dot represents the proportion of the 100 procedures that resulted in a location with the same 90th percentile response time based on the metamodel.

10. Five-Station Results

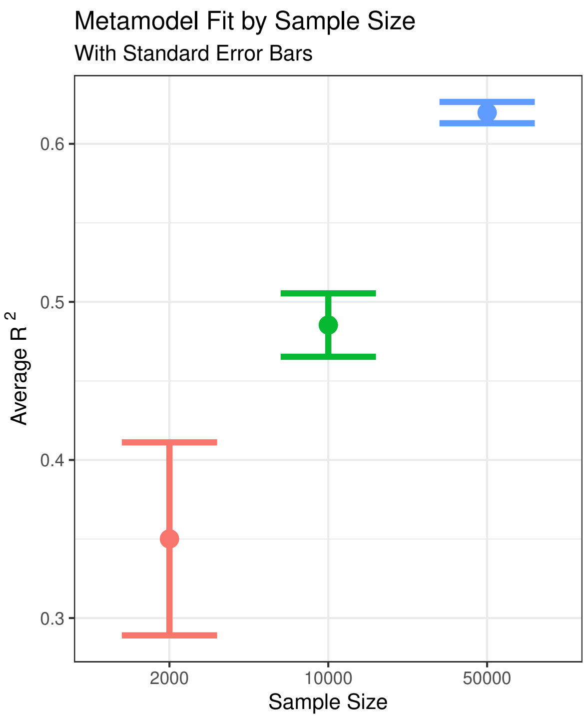
For the more complex settings, the same distributions, factor levels, and regression model described in the previous scenarios were used to create the DES, with parameter values for the gamma distributions listed in Table 5 for the five-station version. A validation of an adjusted DES can be found in Section 3 which compares the response times with actual data.

Time Difference	Factor Levels	Early Morning		Morning		Afternoon		Evening	
		Shape	**Rate**	**Shape**	**Rate**	**Shape**	**Rate**	**Shape**	**Rate**
Call Arrival	**Spring**	1.087	0.011	0.796	0.012	0.952	0.023	1.055	0.021
	Summer	1.080	0.011	0.816	0.012	1.030	0.025	1.026	0.021
	Fall	1.111	0.012	0.769	0.012	1.010	0.025	1.053	0.023
	Winter	1.050	0.011	0.844	0.014	1.010	0.026	1.039	0.021
Assignment to Enroute	**EMS Vehicle**	1.454	0.534	0.782	0.296	0.781	0.409	1.016	0.526
		In City				**Not in City**			
		Shape		**Rate**		**Shape**		**Rate**	
Dispatch to Enroute	**EMS Vehicle**	2.444		1.076		1.895		0.695	
Onscene to Clear	**EMS Vehicle**	1.602		0.026		1.569		0.021	

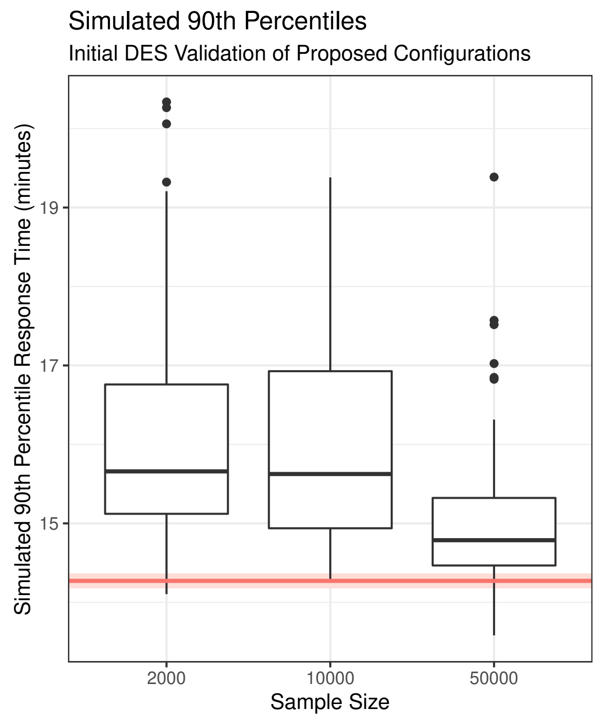
Table 5: Parameter values for the gamma distributions used in the DES for the five-station setting.

Figure 15a shows that the quality of fit improves as a function of sample size, and there is no indication that the rate of improvement is slowing down for the five-station setting. This suggests that given more computational resources, larger sample sizes would result in higher average R^2 values. Figure 15b shows a boxplot of simulated 90th percentile response times for the initial validation of each proposed configuration by sample size, along with a horizontal line representing the average simulated 90th percentile under the current configuration. It indicates that most of the solutions produced by optimizing station locations using the metamodels are, according to the DES, inferior to the current configuration; however, for the 50,000 case, there are a few solutions (represented by the lower whisker on the box-and-whisker plot) that offer a potential improvement. In fact, 13 configurations in the 50,000 case have simulated 90th percentiles below the upper 95% confidence bound of that of the current configuration. For configurations in each sample size that rivaled the 90th percentile of the current configuration, the DES was run four times in total, based on the power analysis, to compute an average. Figure 15c shows the current configuration in red, along with the average and 95% confidence intervals of the simulated 90th percentiles for each of these configurations. None of the proposed configurations based on samples of size 2,000 or 10,000 proved superior to the current configuration, while several configurations based on the sample size of 50,000 had lower average simulated 90th percentiles, though the difference was less than 30 seconds for each.

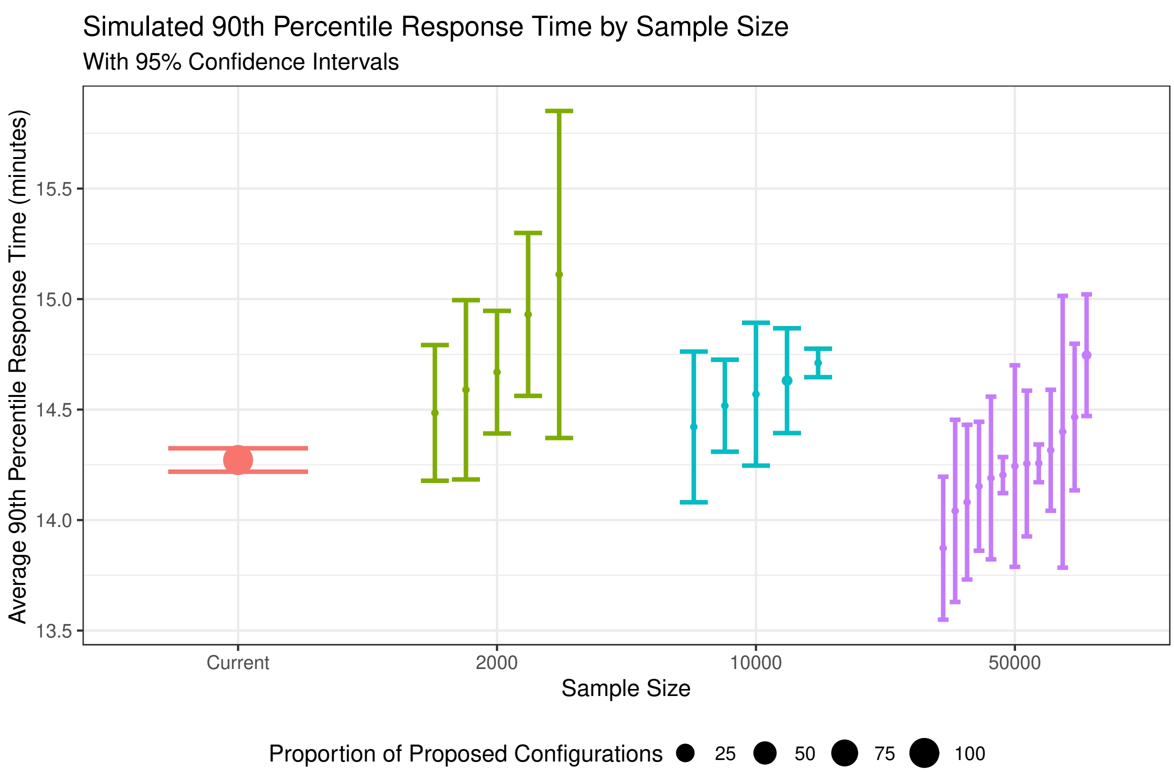
Figure 16 shows the locations of the stations for two of the proposed configurations based on the sample of size 50,000, in comparison to the current configuration in red. The stations shown in green are only slightly shifted from each of the current stations in red, and represent the configuration with the smallest average simulated 90th percentile, with a value of 13.9 minutes compared to 14.3 minutes for the current configuration. The stations in blue represent the configuration with an average simulated 90th percentile of 14.0 minutes, and also resemble the current stations with the exception that there is an additional station to the south of Eveleth rather than in Buhl.



(a)



(b)



(c)

Figure 15: For the five-station setting, (a) metamodel fit by sample size, with standard error bars; (b) boxplot of single DES validation of 100 proposed configurations produced by optimizing the metamodel-predicted 90th percentile response time; and (c) an evaluation of several promising solutions on the DES. The red line in (b) is the DES performance of the current configuration, with 95% confidence band.

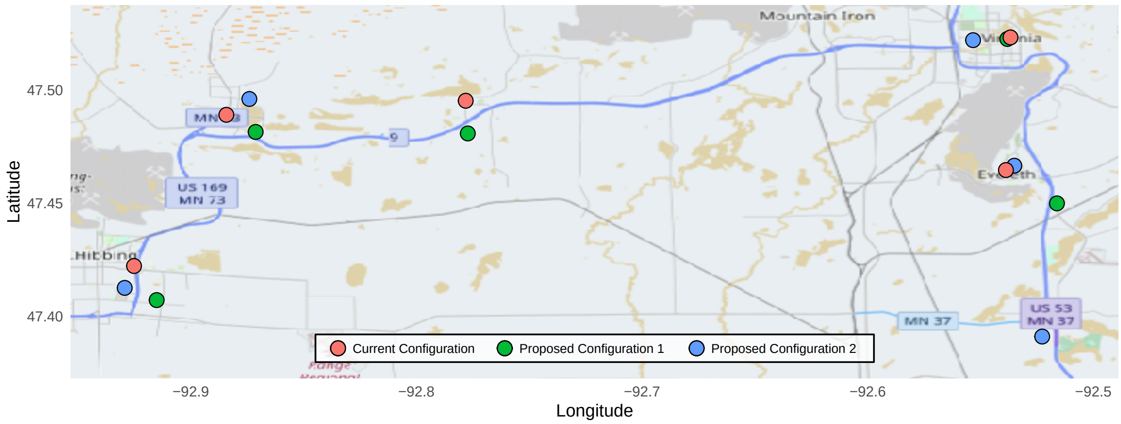


Figure 16: Location of several example proposed configurations for the five-station setting compared to the current configuration in St. Louis County.

11. Details for the Twelve-Station Results (Section 4.3)

The computed parameter values for the gamma distributions in the twelve-station setting are listed in Table 6, which are based on the same factor levels as the other scenarios. All results for the metamodeling, optimization, and validation are presented in Section 4.3 of the article.

Time Difference	**Factor Levels**	**Early Morning**		**Morning**		**Afternoon**		**Evening**	
		Shape	**Rate**	**Shape**	**Rate**	**Shape**	**Rate**	**Shape**	**Rate**
Call Arrival	**Spring**	1.033	0.013	0.835	0.016	0.951	0.028	1.002	0.025
	Summer	0.996	0.013	0.826	0.016	1.038	0.032	0.992	0.026
	Fall	1.134	0.015	0.774	0.015	1.047	0.031	1.029	0.027
	Winter	1.088	0.014	0.837	0.017	1.034	0.032	1.030	0.024
Assignment to Enroute	**EMS Vehicle**	1.282	0.405	0.757	0.245	0.766	0.334	0.964	0.425
		In City				**Not in City**			
		Shape		**Rate**		**Shape**		**Rate**	
Dispatch to Enroute	**EMS Vehicle**	2.468		1.089		1.783		0.644	
Onscene to Clear	**EMS Vehicle**	1.633		0.026		1.766		0.023	

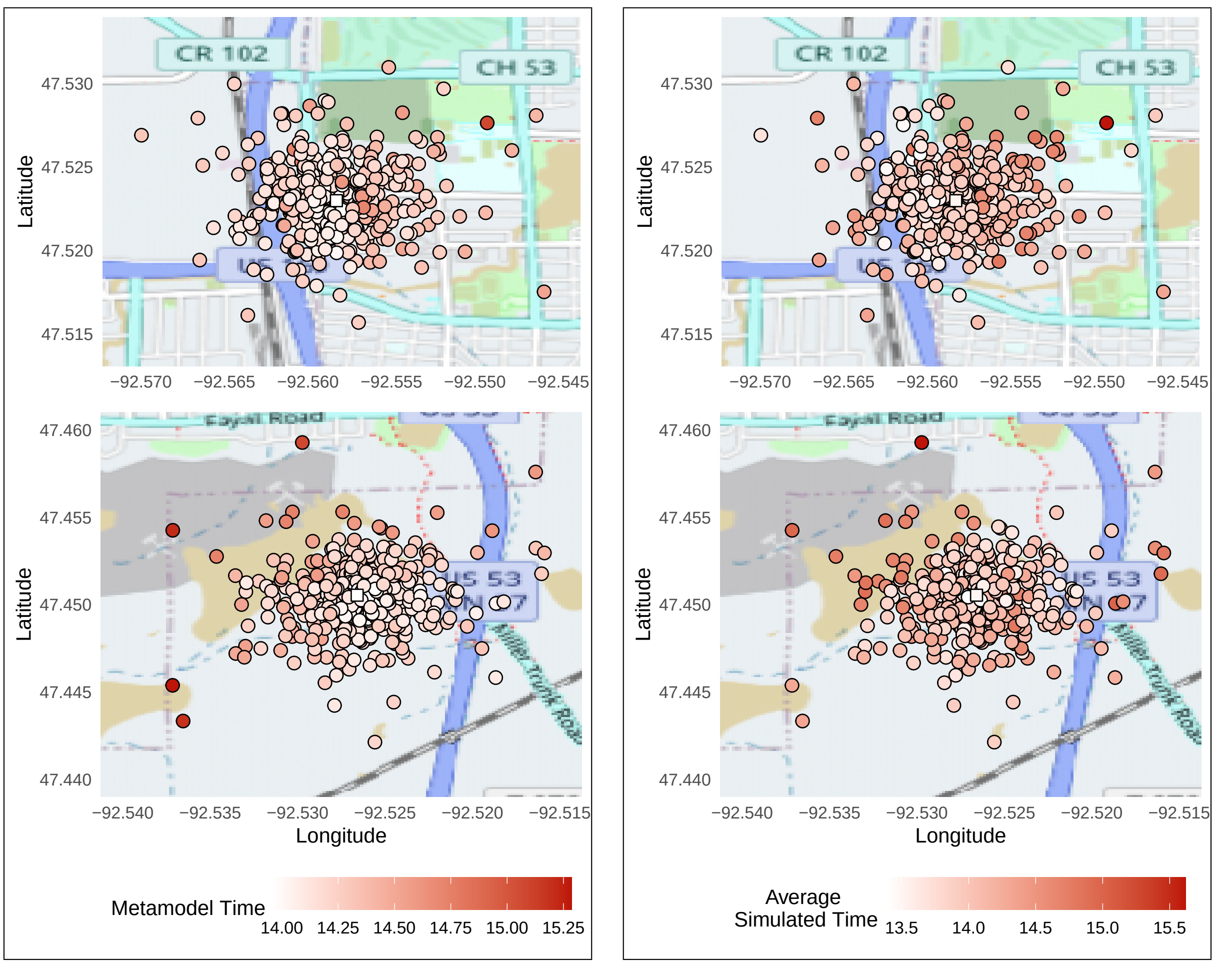
Table 6: For the twelve-station setting, parameter values for the gamma distributions used in the DES.

12. Sensitivity Analyses

In this section, details on two sensitivity analyses are reported.

12.1. Perturbation of selected station locations

A set of 500 perturbed configurations were generated by randomly shifting the station locations by distances (in feet) selected from a Lognormal$(6.5, 0.7)$ distribution and directions (in degrees) selected from a Uniform$(0, 360)$ distribution. The log-normal distributional parameters were selected such that the mode of the distribution was approximately 330 feet, a common distance of a city block, and approximately 90% of the perturbed locations would be within five city blocks, or 1650 feet, of the originally selected locations. Figure 17a zooms in on the two station locations and displays the metamodel-predicted 90th percentiles for each of 500 perturbed configurations. The color of each point represents the 90th percentile, with darker shades indicating longer 90th percentiles. The proposed configuration, shown using squares, has the lowest predicted 90th percentile, with a general, gradual increase in the times as the stations move further from the proposed locations. Similarly, Figure 17b depicts the average simulated 90th percentiles for each perturbation. Here too, the configurations with the smallest mean simulated 90th percentiles tend to be closer to the proposed locations, while those further from the center had higher values. These results are encouraging, implying that the best configuration proposed by these methods does in fact have a low 90th percentile response time, and small perturbations of the locations yield only slightly different results.



(a) (b)

Figure 17: Perturbations of proposed configuration for the two-station setting, color-coded by metamodel-predicted 90th percentile in (a) and average simulated 90th percentile in (b). The top plots represent the northern station; the bottom plots represent the southern station. Squares represent locations of the proposed configuration, and circles represent the perturbations.

12.2. *Perturbation of DES distributional parameters*

Since the results in this paper are dependent on stochastic simulators that were developed using samples of call data, a sensitivity analysis using the one-station setting was performed as well. Specifically, the perturbed α (shape) and β (rate) parameters for the gamma distribution were randomly generated from a multivariate $N([\hat{\alpha} \;\; \hat{\beta}]^T, \mathbf{\Sigma_g})$ distribution, where $\hat{\alpha}$ and $\hat{\beta}$ are the MLEs of the α and β parameters and $\mathbf{\Sigma_g}$ is the corresponding variance-covariance matrix of the gamma parameters, based on the data. Similarly, the perturbed regression coefficients were randomly generated from a multivariate $N(\hat{\boldsymbol{\beta}}, \mathbf{\Sigma_r})$ distribution where $\hat{\boldsymbol{\beta}}$ is a vector of the MLEs of the regression coefficients and $\mathbf{\Sigma_r}$ is the corresponding variance-covariance matrix of these regression parameters. The effects of perturbing the DES parameters are discussed in Section 5.1 of the article.

References

[1] T. Giraud, osrm: Interface Between R and the OpenStreetMap-Based Routing Service OSRM (2020). URL `https://cran.r-project.org/package=osrm`

[2] W. N. Venables, B. D. Ripley, Modern Applied Statistics with S, 4th Edition, Springer, New York, 2002, iSBN 0-387-95457-0.
URL `http://www.stats.ox.ac.uk/pub/MASS4`

[3] D. W. Scott, Multivariate Density Estimation: Theory, Practice, and Visualization (Wiley Series in Probability and Statistics), 1992.

[4] R Core Team, R: A Language and Environment for Statistical Computing (2020).
URL `https://www.r-project.org/`

[5] E. Pebesma, R. Bivand, Classes and methods for spatial data in R, R News 5 2 (2005).
URL `https://cran.r-project.org/doc/Rnews/`

[6] R. S. Bivand, E. Pebesma, V. Gomez-Rubio, Applied spatial data analysis with R, Second edition, Springer, NY, 2013.
URL `https://asdar-book.org/`

[7] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Vol. 103 of Springer Texts in Statistics, Springer New York, New York, NY, 2013. doi:10.1007/978-1-4614-7138-7.
URL `http://link.springer.com/10.1007/978-1-4614-7138-7`

[8] M. Kuhn, caret: Classification and Regression Training (2020).
URL `https://cran.r-project.org/package=caret`

[9] A. Liaw, M. Wiener, Classification and Regression by randomForest, R News 2 (3) (2002) 18–22.
URL `https://cran.r-project.org/doc/Rnews/`

[10] M. Clerc, Standard Particle Swarm Optimisation (2012). URL https://hal.archives-ouvertes.fr/hal-00764996