Beyond Distance and Speed Limit: Estimating Emergency Vehicle Travel Time Using Nationwide Data from Sweden

Lisa Norén
Linköping University
lisa.noren@liu.se

Jonas Larsson
Linköping University
jonas.lar@live.se

Rasmus Ringdahl
Linköping University
rasmus.ringdahl@liu.se

Clas Rydergren
Linköping University
clas.rydergren@liu.se

Krisjanis Steins
Linköping University
krisjanis.steins@liu.se

ABSTRACT

Accurate travel time estimation for emergency vehicles is essential for effective dispatching and response planning. This paper investigates to what extent predictive performance improves when models incorporate variables beyond common inputs such as distance and speed limits. Using nationwide historical GPS data from Swedish emergency vehicles, map-matched trajectories were enriched with multiple variables, including speed limits, speed bumps, precipitation, and vehicle size. An automatic route validation model was developed to identify and remove flawed trajectories. The results show that extending models with additional variables improves estimation accuracy. A compact decision tree was among the best-performing models, improving a baseline model that uses only speed limits as input by 5.61 percentage points in MAPE and 20.26 seconds in RMSE, while more complex models offered limited additional benefits. These improvements reduce uncertainty in dispatch decisions and support more reliable route selection in time-critical emergency response operations.

Keywords

Emergency vehicles, travel time estimation, route validation, variable selection, GPS data

INTRODUCTION

Public safety is reliant on efficient emergency services, as rapid response is critical for minimizing property damage during incidents such as house fires and for saving lives in medical emergencies (Cabral et al., 2018; Zhang et al., 2016). A commonly used performance indicator in this context is response time, defined as the duration from the initial alert to the arrival of service at the incident location. One component of the response time is the travel time, which refers to the duration required for an emergency vehicle to drive from its initial location to the site of the emergency. While distance and speed limit form the basis of many estimation models, their sufficiency for emergency vehicle travel time, especially under varying road and weather conditions, is not as obvious (Cabral et al., 2018; Challands, 2010). Furthermore, it remains unclear to what extent predictive performance is gained by incorporating additional variables, such as road characteristics, weather, and traffic conditions. This paper investigates the role of such variables for estimating emergency vehicle travel time.

Although accurate travel time estimates alone do not directly reduce the response times, they support operational decisions, for example, route selection and vehicle dispatching, which can improve the outcomes of the service. Commercial navigation tools such as Google Maps typically assume compliance with traffic rules, among other things, which makes the tools unsuitable for emergency vehicles driving with blue-lights and sirens (Abid et al., 2024; van Mark et al., 2024). This suggests the need for travel time estimation models for emergency vehicles.

The aim of this paper is to evaluate whether the inclusion of additional variables, beyond distance and speed limit, improves travel time estimation for emergency vehicles, while maintaining model interpretability. Specifically, the following research question is addressed: To what extent can model performance be improved when model inputs are extended beyond distance and speed limit? To investigate this, GPS data from emergency vehicles collected in Sweden between 2010 and 2015 were used, along with contextual information such as unit type (i.e., medical or rescue) and time of day. While the dataset includes several types of emergency vehicles, this study focuses on medical vehicles (e.g., ambulances) and road-based rescue vehicles (e.g., fire trucks), where rescue vehicles is used for the Swedish fire and rescue service (*räddningstjänst*). Throughout this paper, the term emergency vehicles refer collectively to both groups.

The GPS data were map-matched to the Swedish road network to generate analyzable routes, each enriched with variables related to, for example, the vehicle (e.g., size), infrastructure (e.g., speed limits, urban road proportions) and weather (e.g., precipitation). The study is limited by the availability of open-source Swedish road network and historical weather data. Consequently, the results should be interpreted with caution when considering countries with different traffic laws, vehicle fleets, or emergency response structures.

BACKGROUND AND RELATED WORK

Accurate travel time estimation is essential for many location-based services, such as emergency response and public transport (Pham et al., 2024). With today's availability of large-scale historical data, data-driven models can estimate travel times for entire routes by learning from previous trips under similar conditions (Wang et al., 2018). Model performance is typically assessed by comparing model predictions with actual travel times using metrics such as MAPE, RMSE, and R^2. Additionally, the percentage of overestimations and bad prediction rates can be used for evaluations as well (Chicco et al., 2021; Wang et al., 2018; Rahamani & Buzna, 2022). Rahmani & Buzna (2022) argue that overestimations are preferred in this setting, although it is a tradeoff since misallocation of resources increases with excessive overestimation. Wang et al. (2018) define predictions with an absolute error greater than 50% as bad predictions, which, together with other metrics, reveals whether model performance is skewed by outliers. While these evaluation metrics provide a way to compare models, they do not by themselves explain performance gains from broader contextual variables or increased model complexity.

As this study utilized GPS data, route reconstruction was necessary, as raw GPS points represent discrete spatial observations with no information about the whole route. In this study, Fast Map Matching (FMM) was used, a map-matching algorithm based on Hidden Markov Models, selected for its computational efficiency (Yang & Gidófalvi, 2018). However, map-matching is subject to errors, particularly when the vehicle is at a standstill, or when positioning accuracy is degraded, for example due to signal loss. The accuracy and level of detail in the road network data also influence the map-matching results. These errors can introduce unrealistic routes, such as erroneous detours, or unrealistic driving paths through the road network, which in turn affect travel time estimation (Chao et al., 2020). To address this, an automatic route validation step is proposed to remove trajectories with, for example, unrealistic driving behavior, or route geometries not following the original GPS points, from the dataset. Without excluding anomalous routes, map-matching errors could obscure the effects of certain variables by introducing additional noise.

Kolesar (1975) and Kolesar et al. (1975) are among the earliest studies on emergency vehicle travel time estimation. Although it does not involve medical vehicles, they demonstrated how simple linear regression models can be applied, and exemplified models with a high level of interpretability. Kolesar et al. (1975) analyzed how factors such as daylight, darkness, and rush hours influence vehicle speed. More recently, Rahmani and Buzna (2022) evaluated different models using ambulance GPS data to estimate travel speed, but their study was conducted outside Sweden and did not include weather data. Within Sweden, existing research remains sparse and lacks coverage of northern regions, for example, Abid et al. (2024) and Oxenholt and Pétursson (2024). Abid et al. (2024) utilized weather data, but used trip scenarios, not the actual routes, and focused exclusively on ambulance services, while Oxenholt and Pétursson (2024) used a set of predetermined routes in southern Sweden.

This paper addresses these limitations by using a nationwide dataset covering all regions and months of the year and incorporating both medical and rescue services in the same model. It avoids predefined routes and instead map-matches actual GPS trajectories. Furthermore, this paper contributes to the research field by introducing an automatic route validation model, enabling the use of larger datasets in this context, and further explores the impact of incorporating road characteristics, weather conditions, and contextual information simultaneously into travel time models for emergency vehicles in Sweden.

METHOD

This paper is an extension of the work by Larsson and Norén (2025), who developed predictive models for emergency vehicle travel time using a quantitative, data-driven approach. Their workflow consisted of four main steps: (1) selection of relevant data, (2) route generation from GPS point data, (3) application of map-matching using FMM to generate trajectories, and (4) application of automatic route validation. After route validation, features were extracted from various data sources and used as input to the estimation models. Variables included information from the original GPS data, as well as additional data from the Swedish road network and historical weather records. The models were then trained, evaluated, and compared to assess performance and the influence of including different sets of variables. In the following subsections, these parts are explained in more detail.

Data and preprocessing

Historical GPS data from emergency vehicles, road network data, and weather data constituted the three primary data sources. These sources were selected based on their use and relevance in previous studies. Direct traffic flow measurements were not available for this study. Therefore, temporal variables such as time of day and day of week were used as indirect indications, consistent with prior work.

Historical GPS data from emergency vehicles were collected by a company that provides the IT systems used by emergency dispatchers. The dataset covers the years 2010 to 2015 and consists of reports from all vehicles every 20-30 seconds. Each report is represented as a data point with GPS coordinates along with a timestamp and information about the vehicle, such as the unit type (i.e., medical or rescue), vehicle type (e.g., fire truck or water unit for rescue service, and ambulances or organ transport unit for medical service), and response status. Response status indicates the current activity of the vehicle or unit. For example, an ambulance may have a status indicating that it is in route to the patient, transporting a patient to the hospital, assigned to a mission that has not yet started, or temporarily unavailable (e.g., on break). Together with the priority code, this information allows the dataset to be filtered to include only high-priority missions, where vehicles are likely to operate under emergency conditions

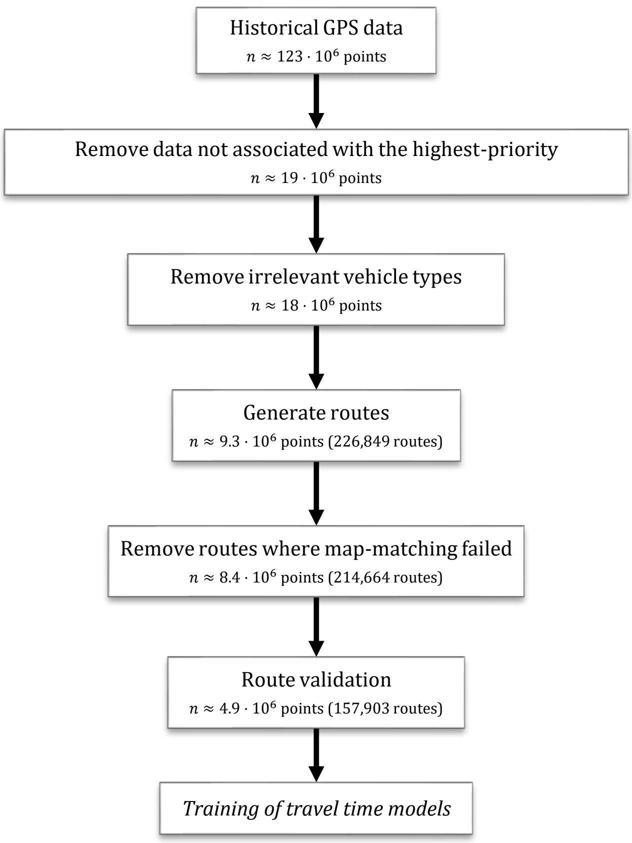


Figure 1. Preprocessing of data with the number of points, n, remaining in the dataset at each step.

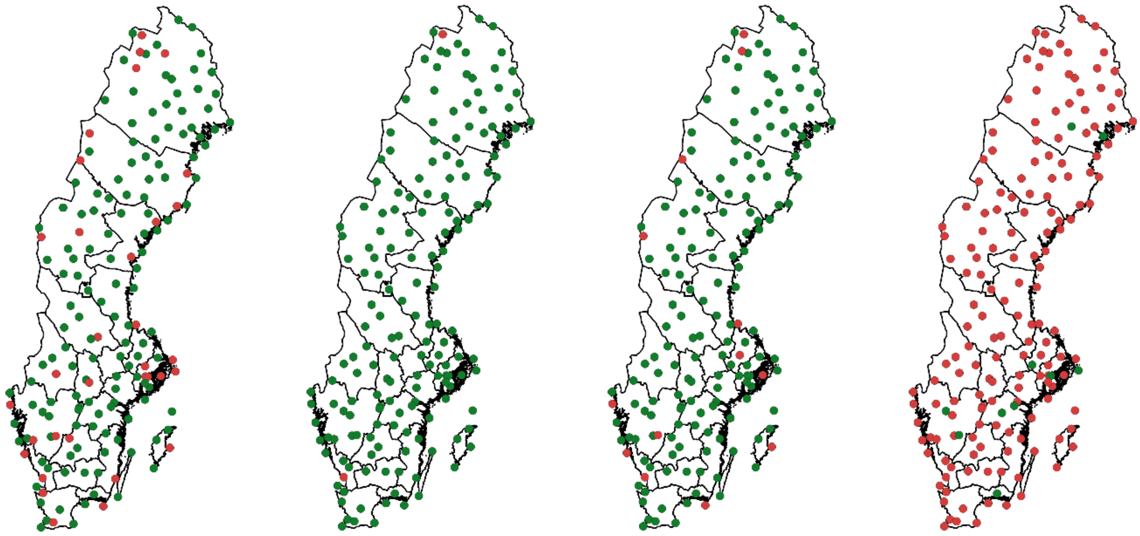
(i.e., with sirens and blue-lights activated). Temporal variables and the ground truth regarding travel times were retrieved through the timestamps, and other contextual variables, such as region of Sweden, were sourced from the GPS data. This data source required the most extensive preprocessing in this study. Vehicles not operating on the road network (e.g., boats and helicopters) were excluded to sustain comparability with road-based infrastructure variables. Figure 1 shows a summary of the preprocessing steps for the historical GPS data, where each data point represents one vehicle report. Route generation, map-matching, and route validation are described separately in the subsequent sections of the paper.

Road network data were extracted from the Swedish National Road Database, NVDB, which is maintained and developed by local authorities and Swedish government agencies (Trafikverket, 2024). The road network data included things such as segment length, speed limits, and speed bumps. By using the historical GPS data from the emergency vehicles, it was possible to match the routes to the road network, where each segment of the network had assigned attributes. It was then possible to determine how often and what proportion of each route a given attribute was present through various calculations.

Weather data were taken from the Swedish Meteorological and Hydrological Institute (SMHI) and included, for example, precipitation (Swedish Meteorological and Hydrological Institute [SMHI], n.d.). The weather data were matched with the routes by using the nearest weather station with the desired attribute to the GPS point in the middle of a route.

Not all available road and weather variables were retrieved. The selection was based on relevance in related work and on sufficient spatial coverage across Sweden. Attributes with sparse or region-specific availability were excluded to avoid geographic bias. Visual inspection confirmed that the retained variables had consistent national coverage, ensuring that observed model performance reflected variable relevance rather than uneven geographic distribution.

Figure 2 illustrates the coverage of chosen weather data variables. The green dots represent the location of weather stations with the specific variable, and red dots mark the location of weather stations without the variable. Out of the four variables displayed in Figure 2, snow depth was the only one determined to have insufficient coverage.



(a) Precipitation **(b) Temperature** **(c) Visibility** **(d) Snow depth**

Figure 2. Spatial coverage of weather data.

Altogether, the GPS data was used to construct routes from consecutive vehicle positions and based on the location, it was possible to add additional variables describing the route. Each route and its variables about the road network, weather, and contextual data were used as input variables in the travel time models.

Route generation and map-matching

To ensure routes reflect meaningful travel patterns, a definition with a set of rules was formulated. A route was defined as a sequence of at least 10 consecutive GPS points belonging to one vehicle identified by its unique vehicle ID, with a total duration longer than 2 minutes, a total distance of greater than 100 meters, and a single response status. A new route was also generated whenever the time between consecutive GPS points exceeded 10 minutes or the distance exceeded 1 kilometer, and when the vehicle changed response status, ensuring that each route contained only one status. An automated rule-based procedure was used to segment GPS points into routes with unique identification numbers.

Initial testing revealed that some GPS trajectories appeared outside of the network, such as through parks or large parking lots, where roads were missing from the NVDB data. Similar issues were observed around hospitals and rescue service stations. Since these locations do not exist in the road network data, the points can never be map-matched correctly. The problem was addressed by removing the first and last three GPS positions of each route, and all positions within 100 meters around hospitals, ambulance stations, and rescue service stations (i.e., fire stations), which also worked to anonymize the data by removing the exact origins and destinations.

Once routes were defined among the GPS points, the routes were map-matched with FMM, and trajectories were created (Yang & Gidófalvi, 2018). FMM also provided outputs regarding map-matched points, such as the shortest path distance between consecutive matched points and the distance between a GPS point and its map-matched point. Additional variables from the road network data and weather data were then added for each route.

Automatic route validation

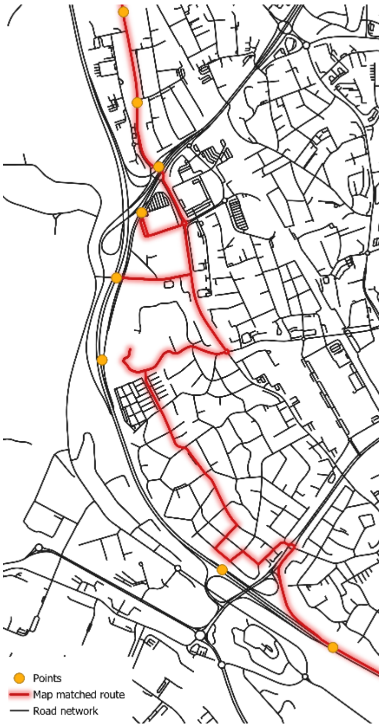
An automatic route validation model was developed to systematically detect and remove poorly map-matched routes. This reduced the reliance on manual inspection and improved the consistency of the data. Figure 3 and Figure 4 show examples of map-matched routes with good and flawed results, respectively. In both figures, the black lines represent the road network, and the red line is the route trajectory made by map-matching the GPS points shown as orange dots. A good map-matched route has the route trajectory and GPS points aligned in a realistic drivable path, where, for example, looping around the same streets is considered unrealistic.

From the original set of routes, a subset was arbitrarily selected and visually inspected manually using MATLAB (MATLAB, 2024). In total, the subset consisted of 5,864 routes, and 70.3% were labeled as good (4,121 routes), meaning that the GPS points and map-matched trajectory looked reasonable, and the remaining 1,743 routes were poorly matched and labeled as bad. The available input variables consisted of, for example, the number of GPS points, changes in speed between consecutive map-matched points, the distance between the GPS point and its map-matched point, the number of consecutive GPS points within a specified radius, and the difference between the shortest path distance for consecutive GPS points and their corresponding map-matched distance. The input variables were based on the emergency vehicle dataset and the map-matching algorithm outputs.

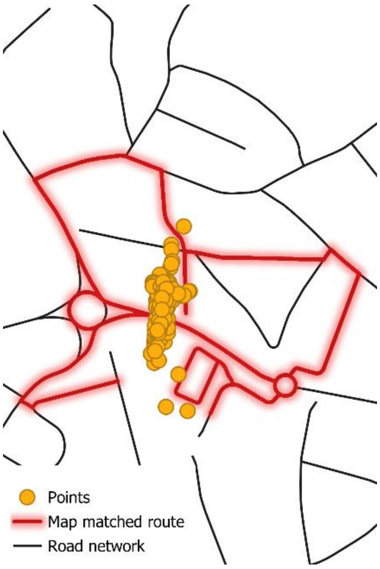
The manually labeled set of 5,864 routes was divided into five parts to enable 5-fold cross-validation when evaluating models, where the false positive rate (the model erroneously predicts the route as good), and overall accuracy were used for evaluation.

(a) Example of a map-matched route with a good result. **(b) Part of a route with a good map-matching result.**

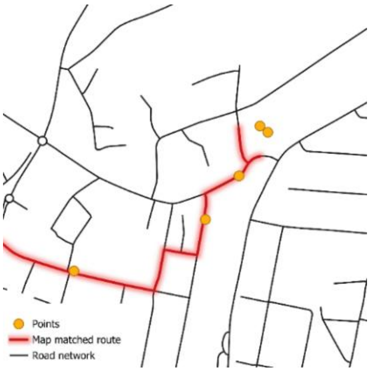
Figure 3. Examples of routes where the GPS points and map-matched trajectory align well.



(a) Trajectory does not follow the GPS points in a realistic driving pattern.



(b) Cluster of GPS points with no obvious route.



(c) GPS points between two road segments, forcing an unrealistic back-and-forth trajectory.

(d) Unnatural detour relative to the GPS points.

Figure 4. Examples of map-matched routes with flawed and bad results.

Automatic route validation results

After testing over 70 models, including decision trees, KNN, lasso, and logistic regression, a generalized linear model (GLM) with a binomial distribution was selected due to its strong performance and interpretability. The evaluated models showed that most underlying models had similar performance, although lasso regression and GLMs were often preferred. These models used input variables that were lasso-derived or manually chosen, and the number of variables in each model ranged from one to all available variables. Initial testing also evaluated scaled versions of the data, both normalization and standardization, but no improvements were identified, and to keep the models interpretable, the unscaled, original data were used.

A generalized linear model (GLM) with a binomial distribution, using seven variables and 21 model coefficients, was chosen as the final model. The accuracy was almost 88% with a probability threshold of 0.5. However, the false positive count was higher than desired, 458 out of 5,864 samples (7.8%). Experiments showed that a probability threshold of 0.90 provided a better balance between accuracy and the false positive rate. This way, precision was favored, prioritizing data quality. Therefore, 0.9 was used as the threshold in the final automatic route validation model, and the accuracy was reduced to 75.3% with 92 false positives out of 5,864 samples (1.6%). Furthermore, additional restriction thresholds were applied, which removed routes with unrealistically high or low speeds.

To assess whether the model introduced bias or not, distribution curves for variables such as unit type, response status, time of day, and total distance were visually inspected and compared before and after applying the model. Results did not show any notable bias, such as removal of one unit type. Although some good routes were misclassified, visual and statistical analyses indicated that the remaining data were representative and sufficient. The process approved 73.6% of the routes, corresponding to 157,903 routes in the dataset out of the 214,664 map-matched routes.

EXPERIMENTS

The travel time modeling experiments were implemented in MATLAB, and model performance was evaluated using standard metrics. The following subsections describe the setup of the experiments, three different approaches used for variable selection, and lastly, other experiments including sensitivity testing and a Google Maps comparison of travel time.

Setup and model development

The travel time estimation models were based on the fundamental relationship between distance, speed, and time, consistent with related work. In this study, route distance was known from the map-matched route trajectories, and the models aimed to estimate the average speed. The estimated travel time was then derived as route distance divided by the predicted speed. This approach was selected because, as noted in the introduction, emergency vehicles often do not follow the speed limits due to their operational nature. Additionally, this approach allowed contextual variables to capture deviations from a baseline model derived from speed limits. Moreover, road characteristics such as roundabouts and urban environments often reduce vehicle speeds below the legal limit, even for regular traffic. All available input variables are described in Table 1 and were extracted from the previously mentioned data sources, except for the variables traffic flow, rush hour, and weekend, which were constructed from combinations of other existing variables.

Categorical variables, for example, region and day of the week, were one-hot encoded, while numerical variables, for example the proportion of routes in urban areas, were treated as continuous. Counter variables, such as the number of speed bumps on a route, were treated as continuous to avoid overfitting due to sparse category representations despite their discrete nature.

The travel time estimation models were developed using statistical and machine learning techniques implemented in MATLAB. All models were trained and evaluated using 5-fold cross-validation, and the best-performing models were also subsequently tested on a final test set. Out of the 157,903 routes available in the dataset, roughly 10% (15,790 routes) were held out as the final test set. Extreme predictions were clipped, where negative travel times were set to zero, and unrealistically large travel times were capped at three times the largest found in the training set. The chosen metrics for evaluation were mean absolute percentage error (MAPE), root mean squared error (RMSE), overestimation rate (OVER), bad prediction rate (BPR), and adjusted R^2 ($AdjR^2$), where most emphasis was placed on MAPE and RMSE. All metrics were found in related work and chosen because of their intuitive connection to the aims and purpose of this paper. Model complexity was also balanced against predictive performance to ensure interpretability and computational efficiency.

Table 1. Available variables in the dataset.

Variable	Type	Description
Average road class	Road	Average road class of the route
Bus street	Road	Proportion of the route on a bus street
Day of week	Contextual	Route carried out during the given weekday
Distance-weighted average speed limit	Road	The average speed limit of the route, calculated by weighting each segment's speed limit by its length relative to the total route [km/h]
Exit entrance	Road	Number of links with an exit or entrance
Hour of day	Contextual	Route carried out during the given hour of day
Industry road	Road	Proportion of the route on an industry road
Main road	Road	Proportion of the route on a main road
Most frequent speed limit	Road	The speed limit with the longest distance of the route [km/h]
Overpass	Road	Proportion of the route on an overpass
Pedestrian street	Road	Proportion of the route on a pedestrian street
Pedestrian zone	Road	Proportion of the route in a pedestrian zone
Precipitation	Weather	Amount of rain approximately at the location of the route [millimeters]
Region	Contextual	Route located in the given administrative region of Sweden
Response status	Contextual	Description of the current activity of the unit
Road class 0, 1, …, 9	Road	Proportion of the route on the given road class (one variable for each road class 0 to 9)
Roundabout	Road	Number of links with roundabouts
Rush hour	Contextual	Route carried out during a rush hour
Southern	Contextual	Route located in the southern half of Sweden, based on a split that ensures equal distribution of GPS data points, before map-matching, between north and south
Speed bumps	Road	Number of links with speed bumps
Temperature	Weather	Temperature approximately at the center of the route [degrees Celsius]
Time-weighted average speed limit	Road	The average speed limit of the route, accounting for varying durations spent on roads with different speeds [km/h]
Total distance	Road	The total length of the road links for the route
Traffic flow	Contextual	Route carried out during morning or afternoon peak
Tunnel	Road	Proportion of the route in a tunnel
Underpass	Road	Proportion of the route in an underpass
Unit type	Contextual	Route driven by a medical or rescue vehicle
Urban area	Road	Proportion of the route in an urban area
Vehicle size	Contextual	Route carried out with a vehicle type of the given size
Visibility	Weather	Estimated visibility approximately at the center of the route [meters]
Weekend	Contextual	Route carried out on a weekend

The type of models explored can be divided into three main phases of experiments: (1) baseline and compact models, (2) correlation-regularized selections, and (3) regularized, stepwise selection and lasso regression models. Phase (1) was used to develop a baseline model by using the time-weighted average speed limit as an input variable. Total distance was also used as a standalone variable and in combination with the speed limit. These models established the marginal benefit of adding other types of variables. The baseline models were extended by using manual variable selection based on domain knowledge and related work. These models contained up to 25 variables. All sets of variable selections were used with linear regression, decision tree, and a pruned version of the decision tree.

Phase (2) addressed multicollinearity, where an automated correlation-based variable filtering was used. Pairwise correlation between all numerical variables was computed, and the variable with the highest variance was kept for each pair. Several different thresholds of correlation were used for the variable filtering. The selected variables, including all binary and categorical variables, were used with linear regression, decision tree, and a pruned version of the decision tree.

Phase (3) explored automated approaches to variable selection and transformation. The models applied lasso-based solutions to identify suitable subsets of variables, and these selections were then used in stepwise regression models that allowed interaction terms. The best-performing models and findings from phase (1) and (2) were used as well.

Sensitivity analysis and comparison to Google Maps

In addition to the experiments, a sensitivity analysis and a comparison to Google Maps estimates were carried out. The sensitivity analysis used a subset of the final test set, where (1) specific regions, and (2) unit types were extracted. The chosen regions were ‘Västerbotten’, ‘Uppsala’, and ‘Halland’, selected arbitrarily to have regions of different sizes and locations in Sweden, while the unit type distinguished between medical and rescue services. The aim of the subset experiment was to reveal insights into potential biases or limitations in the models in terms of prediction accuracy that varies across geographic areas and unit types. Unit type was included as a separate variable because rescue and medical services have typically been modelled separately in related work and because they operate under different operational contexts and vehicle configurations.

The second experiment compared the travel times generated by models from this paper with Google Maps predictions. This experiment used a synthetic dataset constructed with hypothetical emergency locations in the city of Norrköping, represented by random points. To each location, a route was created with FMM to the city’s main hospital and the central fire station. In total, 195 routes were used. Each route was assigned features in the same way as before, and the Google Maps API was used to retrieve the predictions. The time of day for Google Maps estimates matched the temporal variables and the weather. The models used for this experiment utilized all available 157,903 routes from the original dataset in training, and then the models were applied to the synthetic data. The aim of this experiment was to illustrate if emergency vehicle-specific models systematically differ in their estimates from commercial navigation tools, indicating that the proposed models capture travel time patterns distinct from normal traffic and reinforcing the need for dedicated models.

RESULTS

The results section is divided into four parts, where the first part presents the baseline model, followed by the best-performing improved models. The third part describes variable importance and sensitivity analysis, and the final part presents the results of the comparison to Google Maps estimates.

Baseline model

The best-performing baseline model was a decision tree trained using the function *fitrtree* in MATLAB with a maximum of 325 splits, and the time-weighted average speed limit as a single input. Evaluation metrics are found in Table 2. The baseline model was initially compared to a linear regression model and a pruned version of the decision tree. Additional experiments that used total distance as an input, either alone or combined with the speed limit, were carried out as well, but did not yield notable improvements. Figure 5 shows the predicted versus actual travel times in seconds using the chosen baseline model. The red dashed line represents perfect predictions, and each dot represents one route from the final test set. Most dots appear near the red line, but the model tends to underestimate longer travel times.

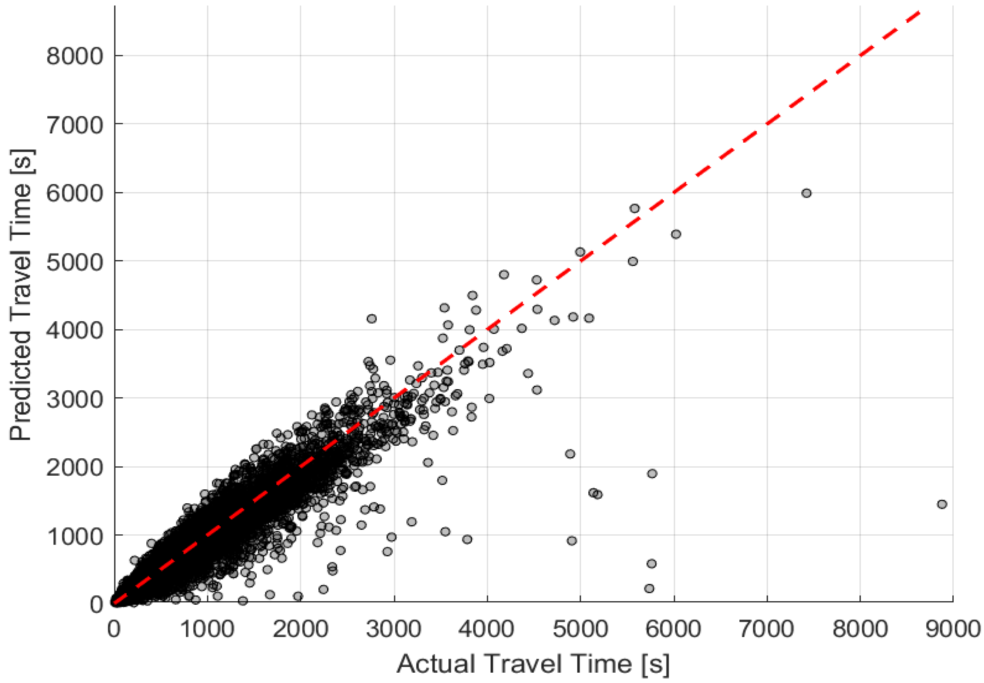


Figure 5. Predicted versus actual travel times for the baseline model.

Improved models

Decision trees performed better than the linear models, in terms of the chosen evaluation metrics, for most variable selections. Similarly, when considering model interpretability and computation time, decision trees were preferred. The travel time estimation errors were within the same range as those reported in related work, and therefore the models presented as improved were determined to be acceptable and sufficient. Most emphasis was given to MAPE and RMSE during the final evaluation and comparison of models because of their frequent use in related work. The number of input variables was also considered to ensure interpretability and therefore models with fewer variables were preferred when models only differed by decimals for MAPE and RMSE.

One of the best-performing models (FM1) across all experiments and model types was a decision tree. MATLAB was set up with the function *fitrtree* and the model argument 'MaxNumSplits' set to 325, controlling the maximum number of splits. The input variables used were total distance, time-weighted average speed limit, hour of day, day of the week, response status, unit type, vehicle size, visibility, percentage of main road, percentage of urban area, number of roundabouts, number of speed bumps, and average road class. The evaluation metrics are presented in Table 2. Figure 6 shows the actual versus predicted travel times using FM1, where the red line indicates perfect prediction and each dot is a route from the final test set. Compared to the baseline, the best-performing models reduced MAPE by about 5.61 percentage points.

A linear-based model (FM2) was found to have similar performance and is presented as an option to FM1. FM2 was constructed using the *stepwiselm* function in MATLAB. The exact model formulation contained logarithmic and square root-transformations of variables and a total of 20 terms. The variables used were total distance, time-weighted average speed limit, traffic flow, and response status. In contrast to FM1, this model does not contain specific route characteristics or weather data, suggesting that models can be improved by other variables. The evaluation metrics are presented in Table 2. Figure 7 shows the actual versus predicted travel times using FM2, where the red line indicates perfect prediction and each dot is a route from the final test set. Compared to the baseline, the best-performing models reduced MAPE by about 4.42 percentage points.

Table 2 also reports evaluation metrics for ten additional models designed to extend the baseline model with groups of variable types. The models R1–R5 include only road-related variables. R1 contains all road variables expressed as proportions (e.g., proportion of the route on a bus street), while R2 includes all road variables that represent counts or binary indicators (e.g., non-proportions). Road class is assigned based on a set of criteria, which indirectly describe other characteristics of the road (Trafikverket, 2020). To investigate that, R3 evaluates a model with the full set of road class proportions, and R4 isolates the average road class as a single aggregated measure. R5 combines all available road-related variables into one model.

The models C1–C4 include contextual variables. C1 contains inputs related to the vehicle (i.e., vehicle size and unit type) and its response status. C2 extends this by adding the original contextual variables, region, hour of day, day of week, and southern, while C3 extends C1 by adding the constructed contextual indicators (i.e., traffic flow, rush hour, and weekend). C4 includes the full set of contextual variables. Lastly, W1 contains all weather-related variables. The decision-tree models consistently outperformed the linear regression models for these variable selections, and therefore only the decision-tree results are presented in Table 2.

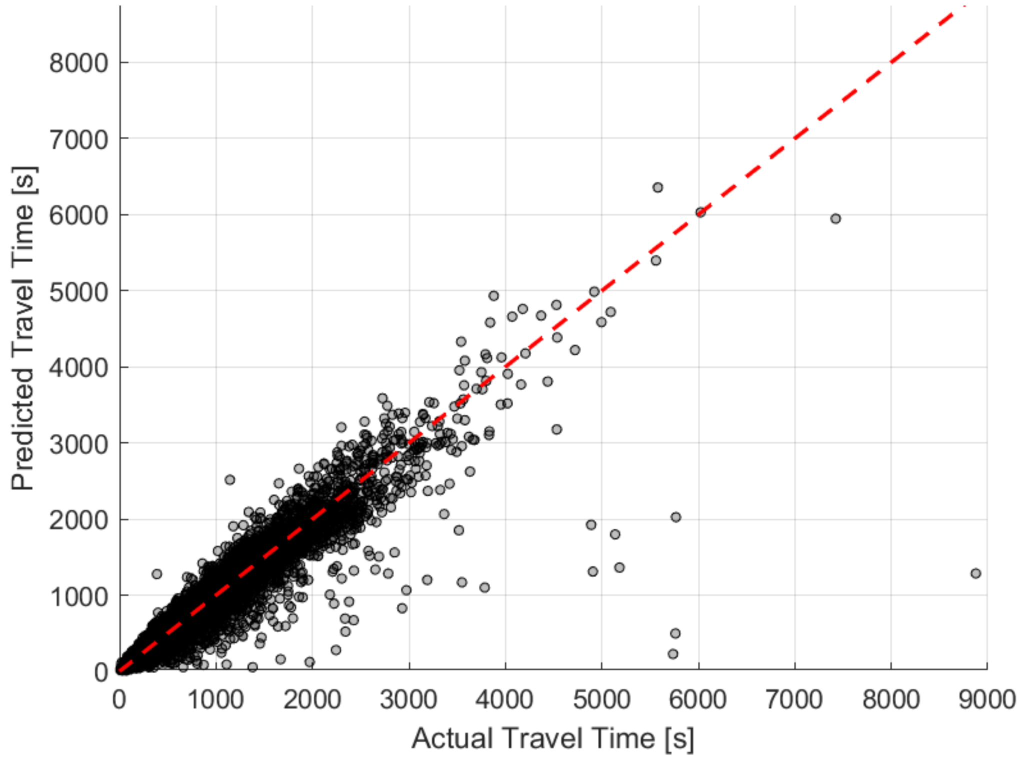


Figure 7. Predicted versus actual travel times for FM1.

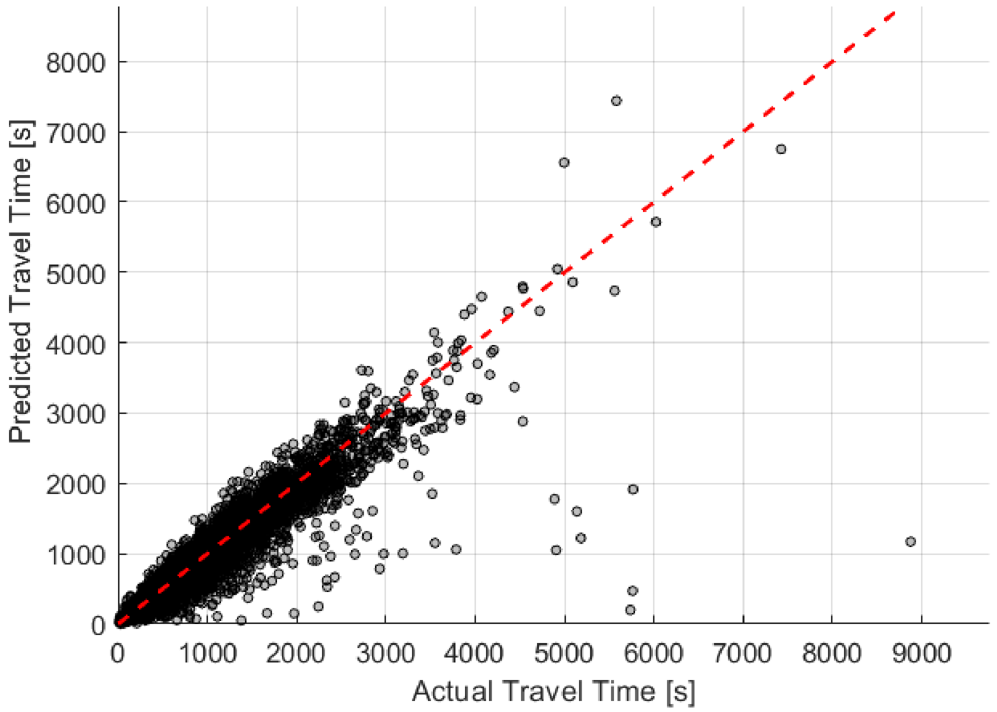


Figure 6. Predicted versus actual travel times for FM2.

Sensitivity analysis and variable importance

Performance of the models varied slightly when regions were isolated in the test set. Results are shown in Table 2. Similarly, test sets with only one unit type, either medical or rescue, showed that it was generally easier to predict the travel time of medical routes. Patterns also showed that travel times for long rescue routes were heavily underestimated. These differences suggest that the relevance of some contextual variables may be region- or unit-specific. Table 2 presents results from the testing with the entire test set and all evaluation metrics for the baseline model, FM1, and FM2, as well as the ten additional models extending the baseline with different variable groups.

Table 2. Evaluation metrics from final testing for all models.

Model	Test set	MAPE[%]	RMSE[s]	OVER[%]	BPR[%]	AdjR^2
Baseline	*All test data*	*21.82*	*202.07*	*47.03*	*6.53*	*0.5409*
Baseline	Västerbotten	19.34	171.45	38.71	3.23	0.5409
Baseline	Uppsala	24.35	168.85	48.12	10.75	0.5409
Baseline	Halland	22.43	164.15	47.76	7.46	0.5409
Baseline	Medical	21.57	186.22	50.21	6.24	0.5409
Baseline	Rescue	22.71	252.17	35.27	7.59	0.5409
FM1	*All test data*	*16.21*	*181.81*	*48.66*	*3.55*	*0.7241*
FM1	Västerbotten	15.09	142.27	34.84	1.29	0.7241
FM1	Uppsala	17.87	153.64	50.00	4.30	0.7241
FM1	Halland	15.79	130.09	57.88	3.15	0.7241
FM1	Medical	15.69	162.47	49.33	3.24	0.7241
FM1	Rescue	18.12	240.18	46.16	4.67	0.7241
FM2	*All test data*	*17.40*	*187.65*	*47.49*	*4.27*	*0.6365*
FM2	Västerbotten	15.94	162.21	35.48	0.65	0.6365
FM2	Uppsala	19.81	153.38	49.73	5.38	0.6365
FM2	Halland	16.98	136.71	54.73	2.65	0.6365
FM2	Medical	16.84	166.67	50.95	3.96	0.6365
FM2	Rescue	19.47	250.45	34.67	5.45	0.6365
R1	*All test data*	20.61	197.77	47.78	5.25	0.6082
R2	*All test data*	21.00	199.11	47.67	5.63	0.5929
R3	*All test data*	*21.19*	*201.48*	*47.45*	*5.72*	*0.5939*
R4	*All test data*	*20.86*	*198.33*	*48.02*	*5.50*	*0.5986*
R5	*All test data*	*20.54*	*196.93*	*48.18*	*5.09*	*0.6097*
C1	*All test data*	*16.57*	*182.69*	*48.08*	*3.97*	*0.6987*
C2	*All test data*	*16.38*	*181.84*	*48.29*	*3.81*	*0.7154*
C3	*All test data*	*16.57*	*182.19*	*47.94*	*3.96*	*0.6992*
C4	*All test data*	*16.37*	*181.82*	*48.27*	*3.83*	*0.7154*
W1	*All test data*	21.60	202.17	47.44	6.05	0.5760

The additional models demonstrate that the baseline model is notably improved when contextual variables are included. Analysis of variable importance for the decision trees showed that the time-weighted average speed limit variable, followed by total distance and response status, dominated early splits and the importance score for most models, not only for FM1, although the results varied depending on all present variables. However, weather variables had a limited impact across all evaluated models. Figure 8 shows importance scores for the final model, FM1. These were computed using the MATLAB function *predictorImportance*, which estimates how much each predictor reduces the mean squared error when used for splitting in the decision tree.

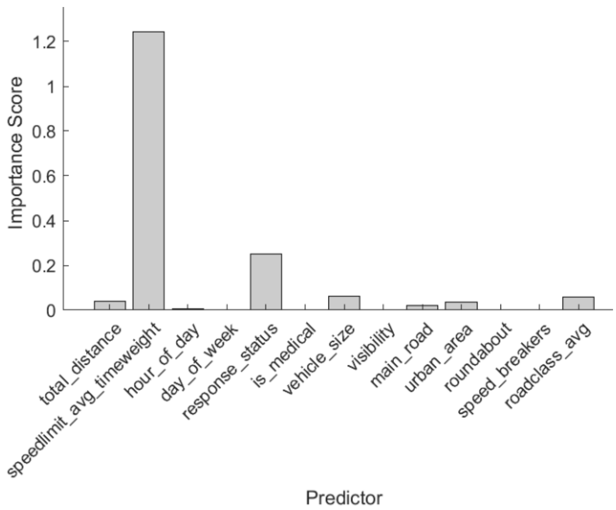


Figure 8. Predictor importance score for variables used in FM1.

Table 3 shows the number of times the variables were used for splitting in the first 50 nodes in the decision tree. These variables are shown in Figure 8 with relatively high importance scores as well. Altogether, this shows how these variables affect the structure of the decision tree and explain major parts of the variability in the travel time.

Table 3. Top five variables used for splitting in the first 50 nodes in FM1.

Variable	Count
Time-weighted average speed limit	17
Vehicle size	8
Proportion of the route on main roads	6
Total distance	6
Response status	6

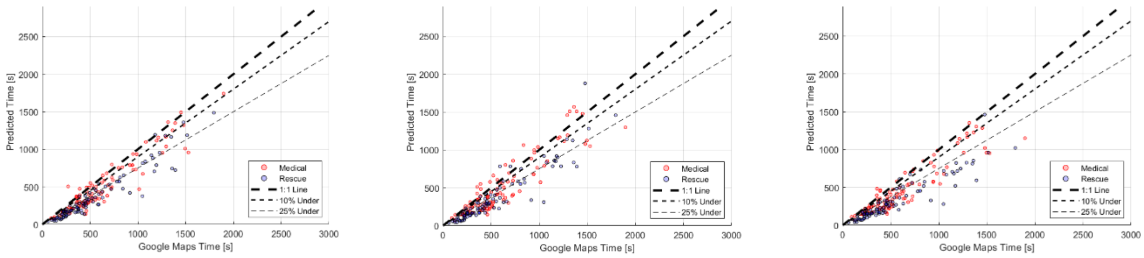
Comparison of model to Google Maps estimates

The synthetic routes created for the Google Maps comparison showed that the final models frequently predicted shorter travel times. Table 4 summarizes the absolute and percentage differences between estimations made by the models and Google Maps. The columns show the minimum, maximum, and average absolute difference in seconds and percentages. The final column shows the percentage of routes where the model prediction was equal to or higher than Google Maps' estimated travel time (≥ 100%), and less than 90% and 75% of the Google Maps time, respectively. Depending on the final model used, over 77% of the routes from the synthetic test set were estimated to be ≤ 90% of Google's time, and over 58% of the routes were shorter than 75% of Google's time. For longer routes, it was possible to distinguish between medical and rescue vehicles, where rescue vehicles showed a stronger trend of having shorter estimations compared to Google Maps. Note that the results of this experiment are indicative, because there is no ground truth for the routes, and the comparison is intended to illustrate behavioral differences rather than benchmark absolute accuracy. It is also unknown which variables are used and how the estimation is computed in Google Maps.

Table 4. Travel time estimations using the final models compared to Google Maps estimates.

Model	Absolute Difference [Min, Max, Avg] (s)	Absolute Percentage Difference [Min, Max, Avg] (%)	Higher or Lower [≥ 100%, <90%, <75%]
Baseline	[1.90, 671.42, 152.09]	[0.48, 86.77, 29.79]	[7.69, 81.54, 57.95]
FM1	[0.52, 737.90, 163.30]	[0.13, 70.41, 31.10]	[13.33, 77.44, 57.44]
FM2	[0.09, 799.54, 187.80]	[0.01, 73.16, 33.84]	[5.64, 84.10, 67.69]

Figure 9 compares the model-predicted travel times to Google Maps estimates for synthetic test routes. Each dot represents a route, with red and blue indicating medical and rescue units, respectively. The thick black dashed line shows the 1:1 relationship, representing equal estimates, while the thinner lines indicate model predictions that are 10% and 25% lower than Google Maps estimates.



(a) Baseline model **(b) FM1** **(c) FM2**

Figure 9. Predicted travel time by the final model versus Google Maps estimates.

DISCUSSION

The results showed that travel time estimations for emergency vehicles can be improved by incorporating contextual variables and extending the model beyond distance and speed limit. Travel distance and time-weighted average speed limit still dominate the models if they are present and show strong predictive contributions, which is consistent with previous studies. However, adding a small set of additional variables such as response status, vehicle size, road class, and proportion of the route in urban areas significantly improved model performance. The improvement was achieved without substantially increasing model complexity, which suggests that compact and interpretable models can capture most of the predictive value. Furthermore, the automatic route validation model ensured data quality by removing flawed trajectories (routes) after the map-matching. The validity of the results was strengthened because it ensured that the improvements observed reflected meaningful variables, not data artifacts.

The overestimation rate presented in Table 2 provides insight into precision, and while overestimations are preferred to be conservative and realistic in planning, they may lead to inefficient resource use. The models do not show significant differences, although FM1 is closer to 50% than the baseline and FM2, reflecting a slightly better balance. The bad prediction rate highlights cases where estimates significantly deviate from the actual travel time. Both FM1 and FM2 showed better BPR than the baseline model, suggesting that the robustness of the models improved. The adjusted R^2 values improved with FM1 and FM2 compared to the baseline model. Furthermore, FM1 showed better results than FM2, suggesting that the model explains a substantially greater proportion of variance even when accounting for model complexity.

Although the best-performing decision tree model, FM1, has a larger set of input variables compared to the baseline model, both use the same underlying model structure and model arguments. FM2 has a smaller set of input variables than FM1 and differs in its model structure but has similar performance. However, FM2 requires nonlinear elements and several variable transformations, which leads to a loss of interpretability.

From a conceptual perspective, weather conditions are expected to influence driving behavior for any type of vehicle. However, weather-related variables, for example visibility, showed limited impact on the overall model accuracy according to the variable importance and split counts in Figure 8 and Table 3. This finding may have several explanations. First, weather conditions may have a relatively modest effect on emergency driving behavior compared with road characteristics and contextual variables. Second, the spatial distribution of weather stations may have been insufficient to capture localized or short-lived extreme conditions, such as heavy rain showers, icy patches, or snowstorms, that are likely to affect driving speed and road accessibility. In addition, the variable encoding and temporal aggregation used in the study, for example visibility measured in meters and precipitation measured in millimeters, may have further reduced the observable impact of weather-related variables. Other potentially relevant indicators, such as road slipperiness, were not available in the dataset. A more thorough analysis of weather effects was not possible within the scope of the present dataset. In particular, the observation period may have contained relatively few days with extreme weather conditions (e.g., heavy snow or very poor visibility), which limited the statistical power to identify stronger weather-related effects.

Road-related variables, including the proportion of main roads and average road class, contributed to improved prediction accuracy, although their effect was moderate. These variables likely capture the driving environment that influences how emergency vehicles maneuver through traffic. Similarly, contextual variables, such as response status and vehicle size, indicate operational constraints that are important during emergency responses. Together, these attributes connect the raw trajectory data, i.e., routes, to real-world driving behavior, which reinforces the value of domain awareness in variable selection.

Nonlinear models, such as decision trees, outperformed linear models when interaction terms and stepwise regression were not in use. This suggests that nonlinearities exist but can often be approximated with a linear-structured model, for example FM2, which could be preferable for real-time use, because of its computational efficiency when retrieving variable values. The sensitivity analysis further indicated that model performance may improve by dividing the models based on unit type (i.e., medical and rescue vehicles). The comparison to Google Maps illustrated that travel time models for emergency vehicles should be developed separately from normal traffic.

While the modeling and analysis provide valuable insights into the factors affecting travel times and their variation across unit types, the resulting model could also support accessibility analyses, resource optimization, and the simulation of emergency response systems. In addition, it could be integrated into existing dispatcher software for route calculation or decision-support purposes. To use the findings from this paper in real-life applications, it is essential to benchmark the presented model results against those currently used by emergency dispatchers and services. There was no such baseline available during the study but establishing one would clarify the practical relevance of the results and indicate whether the observed accuracy gains justify further development. The time-critical nature of many emergency conditions underscores the operational relevance of accurate travel time estimation. Survival rates decrease rapidly for several life-threatening situations, such as major bleeding or cardiac arrest, with each minute without appropriate intervention. In such contexts, even small differences in travel time influence outcomes. More accurate travel time estimates therefore reduce uncertainty in dispatch decisions and increase the likelihood that the unit capable of arriving fastest is selected, which is essential when time is critical.

CONCLUSION

This study developed and evaluated travel time estimation models for emergency vehicles using a nationwide dataset from Sweden. The results show that models extended with contextual variables, such as response status, and road characteristics, such as the proportion of the route on main roads, outperformed models relying solely on distance and speed limits by measurably improving MAPE and RMSE. The best-performing model, FM1, improved MAPE by 5.61 percentage points and RMSE by 20.26 seconds relative to the baseline model, with additional improvements in the bad prediction rate and adjusted R^2. Weather-related variables, such as precipitation and visibility, contributed less to model performance than road-related and contextual variables, likely due to aggregation effects or weaker connections to emergency driving behavior. The automatic route validation model ensured that observed model improvements were based on meaningful data rather than poor map-matching, strengthening the reliability of the results. Decision trees consistently delivered strong performance, and stepwise linear models with transformed variables also performed well, demonstrating that accuracy gains can be achieved without sacrificing interpretability or computational efficiency.

In conclusion, travel time estimation models extended beyond the common inputs of travel distance and speed limits improve estimation accuracy in terms of reduced MAPE and RMSE by at least 5 percentage points and 20 seconds, respectively. Contextual variables are the strongest contributors to the improvements and the models remained interpretable and computationally efficient, making them suitable for real-time applications. Accurate travel time estimates reduce uncertainty in dispatch decisions, enabling the identification of faster routes and increasing the likelihood that the unit closest in terms of travel time is selected, which in turn supports more efficient resource utilization. Although the individual improvements may appear small in absolute terms, they accumulate across the large number of emergency responses handled over time, resulting in meaningful operational effects where seconds can influence outcomes.

FUTURE WORK

Given the strong performance of decision trees, future work could explore ensemble methods. To enhance the automatic route validation model, a semi-supervised learning approach could reduce the manual effort required for labeling while increasing efficiency. For the travel time model, improving the encoding of categorical variables and incorporating additional features such as elevation, better indicators for extreme weather or detailed traffic flow indicators could further boost performance. Developing separate models for urban and rural areas may also improve accuracy due to differences in infrastructure and traffic patterns. Future work should also benchmark the models against currently used dispatcher practices to quantify how the demonstrated performance gains translate into practical improvements.

ACKNOWLEDGMENTS

This research was carried out within the Center for Advanced Research in Emergency Response (CARER) with funding from Linköping University and Swedish Civil Contingencies Agency (MSB).

REFERENCES

Abid, M. A., Lorig, F., Holmgren, J., & Petersson, J. (2024). Ambulance Travel Time Estimation using Spatiotemporal Data. *Procedia Computer Science, 238*, 265-272. doi:10.1016/j.procs.2024.06.024

Cabral, E. L. D. S., Castro, W. R. S., Florentino, D. R. M., Viana, D.A., Costa Junior, J. F. D., Souza, R. P., Rêgo, A. C. M., Araújo-Filho, I., & Medeiros, A. C. (2018). Response time in the emergency services. Systematic review. *Acta Cir Bras, 33*(12), 1110-1121. doi:10.1590/s0102-865020180120000009

Challands, N. (2010). The Relationships Between Fire Service Response Time and Fire Outcomes. *Fire Technol, 46*, 665–676. doi:10.1007/s10694-009-0111-y

Chao, P., Xu, Y., Hua, W., & Zhou, X. (2020). A Survey on Map-Matching Algorithms. In R. Borovica-Gajic, J. Qi, & W. Wang, *Databases Theory and Applications. ADC 2020. Lecture Notes in Computer Science(), vol 12008.* Cham: Springer. doi:10.1007/978-3-030-39469-1_10

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science, 7:e623*. doi:10.7717/peerj-cs.623

Kolesar, P. (1975). A Model for Predicting Average Fire Engine Travel Times. *Operations Research, 23*(4), 603-613. Retrieved from https://www.jstor.org/stable/169845

Kolesar, P., Walker, W., & Hausner, J. (1975). Determining the Relation between Fire Engine Travel Times and Travel Distances in New York City. *Operations Research, 23*(4), 614-627. doi:10.1287/opre.23.4.614

Larsson, J., & Norén, L. (2025). *Estimating emergency vehicle travel time.* Master's thesis, Linköping University, Department of Science and Technology (ITN). Retrieved from urn:nbn:se:liu:diva-215984

MATLAB. (2024). version 24.2 (R2024b). Natick, Massachusetts: The MathWorks Inc.

Oxenholt, H., & Pétursson, I. (2024). *Improving estimation of ambulance travel time by pre-processing and analyzing historic transports and weather.* Bachelor's thesis, Malmö University, Department of Computer Science and Media Technology. Retrieved from https://mau.diva-portal.org/smash/get/diva2:1875926/FULLTEXT02.pdf

Pham, T. S., Nistor, M. S., Cao, L., Gerschberger, M., & Moll, M. (2024). Machine Learning in Vehicle Travel Time Estimation: A Brief Technological Perspective and Review. *Proceedings of the 57th Hawaii International Conference on System Sciences*, (pp. 1409-1414). doi:10.24251/HICSS.2024.175

Rahmani, S., & Buzna, Ľ. (2022). GPS-based data driven modeling of ambulance travel times: The case of Žilina region. *2022 IEEE International Conference on E-health Networking, Application & Services (HealthCom)*, (pp. 234–239). doi:10.1109/HealthCom54947.2022.9982741

Swedish Meteorological and Hydrological Institute (SMHI). (n.d.). *Introduction*. Retrieved 2025-03-18, from https://opendata.smhi.se/metobs/introduction

Trafikverket. (2020). *Dataproduktspecifikation – Funktionell vägklass, Version: 9.0.* Trafikverket. Retrieved from https://bransch.trafikverket.se/TrvSeFiler/Dataproduktspecifikationer/V%C3%A4gdataprodukter/DPS_E-G/1010funktionell_vagklass.pdf

Trafikverket. (2024). *About nvdb*. Retrieved 2025-03-18, from https://www.nvdb.se/sv/about-nvdb/

van Mark, A., Hallstein, T., Holzgreve, F., Groneberg, D. A., & Ohlendorf, D. (2024). How do different navigation systems affect emergency response time? A prospective simulation study. *BMJ Open*, Article 14:e079094. doi:10.1136/bmjopen-2023-079094

Wang, Z., Fu, K., & Ye, J. (2018). Learning to Estimate the Travel Time. *KDD '18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, (pp. 858 - 866). doi:10.1145/3219819.3219900

Yang, C., & Gidófalvi, G. (2018). Fast map matching, an algorithm integrating hidden Markov model with precomputation. *International Journal of Geographical Information Science, 32*(3), 547-570. doi:1080/13658816.2017.1400548

Zhang, Z., He, Q., Gou, J., & Li, X. (2016). Performance measure for reliable travel time of emergency vehicles. *Transportation Research Part C: Emerging Technologies, 65*, 97-110. doi:10.1016/j.trc.2016.01.015