American Journal of Emergency Medicine 37 (2019) 1633–1642

Contents lists available at ScienceDirect

American Journal of Emergency Medicine

journal homepage: www.elsevier.com/locate/ajem

Original Contribution

# Using spatial regression methods to evaluate rural emergency medical services (EMS)

Zhaoxiang He [a], Xiao Qin [a,*], Ralph Renger, Ph.D. [b], Eric Souvannasacd, MS [c]

[a] Department of Civil and Environmental Engineering, University of Wisconsin-Milwaukee, Milwaukee, WI 53201-0784, United States of America
[b] Center for Rural Health Evaluation, University of North Dakota, 250 Centennial Drive Stop 8138, Grand Forks, ND 58202-8138, United States of America
[c] Center for Rural Health, University of North Dakota, School of Medicine & Health Sciences, 1301 N. Columbia Road, Stop 9037, Grand Forks, ND 58202-9037, United States of America

ARTICLE INFO

*Article history:*
Received 4 June 2018
Received in revised form 5 October 2018
Accepted 20 November 2018

*Keywords:*
Emergency medical service (EMS)
Response time
Spatial lag model
Spatial error model
Geographically weighted regression

ABSTRACT

Emergency Medical Services (EMS) are acute services provided outside of the hospital. EMS are crucial in rural environments where hospitals are often far away and difficult to access. Establishing EMS performance measures is critical in improving a rural community's access to these services and eliminating systemic inequalities. However, an absence of data leads to challenges in developing objective and quantifiable service metrics. EMS data are regularly collected through the National EMS Information System (NEMSIS), yet the manner of data collection and quality of data vary across agencies. Moreover, the amount and complexity of information makes data analyses difficult, subsequently effecting EMS leaderships' ability to identify improvement needs.

This study used NEMSIS data to exemplify approaches for establishing two data-driven performance measures. The measures used in this study – timely service and service coverage – are both dependent on the mobility and accessibility of the EMS transportation network. Two types of spatial models: the spatial econometric model and geographically weighted regression (GWR) model, were developed and then compared to the linear regression model to help identify response time factors. GWR performed best in terms of goodness-of-fit statistics and was chosen to help understand how factors (e.g., weather, transportation) impact the timely provision of EMS in rural areas. The GWR results provided additional insights through the particular spatial patterns of the coefficient estimates and their statistical significance to EMS practitioner for their references to reduce local response times.


## 1. Introduction

Rural emergency medical services (EMS) face unique challenges related to limited resources, sparsely distributed populations, and long transport times [1]. There are real-world impacts of these challenges. For example, the fatality rate for motor vehicle crashes in rural areas is considerably higher when compared with urban areas [2,3]. Each year >30,000 people lose their lives on United States roads, approximately 70% of those fatalities occur on rural roads. According to the National Highway and Transportation Safety Administration (NHTSA), “Delay in delivering emergency medical services is one of the factors contributing to the disproportionately high fatality rate for rural crash victims” [4]. Addressing rural EMS availability and service performance is key to improving emergent 911 call/incident survival rates. EMS systems' quality improvement initiatives are based on established performance measures. But, rural EMS quality improvement endeavors are hindered by a lack of appropriate evaluation methods and performance metrics [5-7].

Improving access to rural EMS is often challenging due to lack of resources, personnel, infrastructure, and long transport times [8]. This context is often compounded because the operational structure of EMS systems varies across communities. For example, EMS can be located within a fire department, hospital, or a stand-alone agency, governed by municipality or county, or even run privately. Further, rural EMS services can be volunteer or professional, municipal or private, air or ground, large or small. The vast differences across EMS services creates challenges for capturing system-wide information using consistent performance metrics. Having similar metrics would allow rural EMS services to aggregate data across services. This is especially important since any rural service typically has an insufficient number of time urgent cases (e.g., cardiac arrest, STEMI, etc.) upon which to base improvements to operating procedures. In recent years, NEMSIS – the national repository used to store all United States' EMS data – has helped mitigate some of these challenges [9], but a lack of compliance across services still yields unreliable data.

* Corresponding author.
*E-mail addresses:* he22@uwm.edu (Z. He), qinx@uwm.edu (X. Qin), ralph.renger@med.und.edu (R. Renger), eric.souvannasacd@med.und.edu (E. Souvannasacd).

https://doi.org/10.1016/j.ajem.2018.11.029
0735-6757/

Despite the challenges, there are some promising data being collected to evaluate rural EMS performance. For example, NHTSA developed 35 EMS performance measures for local EMS services, including both time-based (e.g., mean emergency patient response interval) and outcome-based (e.g., EMS cardiac arrest survival rate, emergency department discharge) variables. The local service chooses which variables to use to evaluate its service performance based on its situation [10].

Although a patient's outcome depends on many factors, the time required for an EMS unit to arrive at the scene (or response time) and the time required for a patient to receive definitive care (or overall response time) plays a key role in their survival. The Centers for Disease Control and Prevention (CDC) reports that when trauma victims receive definitive care at a level I trauma center, the mortality risk is reduced by 25% [11]. However, crashes or other severe incidents occurring in rural areas can happen far away from a level I trauma centers, meaning quick response and fast transport to the receiving agencies are difficult to achieve. Therefore among all performance measures, response time is considered one of the most important to improve emergent incident survival rates [12]. Thus, to improve overall rural EMS patient outcomes, response time data must be analyzed to identify factors effecting EMS performance.

The linear regression model is popular, for identifying the statistical relationship in which the time-related performance is the dependent variable and contributing factors are the independent variables. However, because this model was not designed to handle spatial dependence, results can provide inaccurate coefficient estimates and unreliable statistical inferences if spatial dependence exists. Spatial dependence on the model's residuals and spatial heterogeneity are the two emerging issues when a statistical regression is conducted on spatial data. Anselin developed the spatial lag model and spatial error model to solve the issue of spatial autocorrelation [13]. A spatial lag model is applied when a spatial structure exists in the response variable. A spatial error model is applied when a spatial structure exists in the residuals [13]. Such spatial models are widely used in economics [14,15] and are applied to study transportation issues [16,17]. One study found a spatial lag model and spatial error model performed better than a linear regression model when investigating the impact of accessibility and weather on emergency unit reaction times [18].

When the variable coefficients vary across space or are weighted geographically, they are best addressed by the geographically weighted regression (GWR) model [19]. The GWR model was used to account for spatial heterogeneity in previous studies [20-24]. Zhao et al. used the GWR model to estimate annual average daily traffic and found the GWR model had a better goodness-of-fit than the generalized linear model [20]. Du et al. used the GWR model to capture the relationship between transport accessibility and land value [21]. Li et al. used the geographically weighted Poisson regression for county-level crash modeling in California [22].

This paper is an empirical study utilizing NEMSIS data to establish and highlight plausible performance metrics for EMS in rural areas. The study aims to measure timely service and service coverage, and to use spatial statistical methods (e.g., spatial econometric models and GWR model) to identify factors significantly affecting EMS performance. The findings of the study are instrumental in assisting decision-makers with resource optimization, strategic planning for rural EMS, and tactical placement or relocation of ambulance services.

## 2. NEMSIS dataset & service-based performance measures

This study used the 2013 NEMSIS dataset obtained from the South Dakota EMS office, and South Dakota roadway information as provided by the South Dakota Department of Transportation. According to NEMSIS, 36,198 emergency (911) calls were answered by 109 South Dakota EMS stations. Detailed descriptions regarding EMS data attributes and data quality assurance/quality control (QA/QC) are omitted for brevity, as more information is available from Qin et al. [25]. After removing invalid information (e.g., extremely high values of response time), 13,041 emergency records with valid response times were used to evaluate service performance (i.e. response time). Response time in this study is the time interval between two consecutive time-stamped events: the length of time between when the responding unit starts moving to the time the responding unit physically arrives at the scene. Note that this time interval does not include the length of time that passes during the 911 call or the time it takes to dispatch the first responders.

The quantitative assessment provides the measure of coverage area for each EMS service, in square miles, that can be served by an EMS service within an acceptable time interval. Based on the requirement for South Dakota EMS, a 15-minute threshold value was used for response time [26]. One caveat for this network-level analysis is that the area covered by 15-minute travel time was calculated by the estimated travel speed and for the areas with unknown speed limit, the average response speed of 35 mph was used. Fig. 1 presents a 15-minute benchmark for response time in which the coverage area was generated by the road network and network analyst toolbox in ArcGIS [27]. Location coordinate information of each incident was retrieved from Google Maps APIs.

As shown in Fig. 1, only a portion of the state was covered by EMS within the 15-minute travel time. Counties with a more sparsely distributed population (e.g., Shannon County) have more points uncovered by EMS. Coverage ratio, or the number of 911 calls within the 15-minute coverage area for all EMS stations over the total call volume, was calculated as 71%. This relatively low value is due to sparsely distributed demands, which is common in rural areas.

Fig. 2 presents two performance indexes: the service coverage ratio index and the service timeliness index. The service coverage ratio index is defined as the number of cases within the 15-minute coverage area over the total number of calls responded to by an EMS service. The service timeliness index is defined as the percentage of cases with an actual response time of no more than 15-min within the 15-minute coverage area for each station.

Fig. 3 offers a closer view of EMS performance. Among 13,041 cases with complete information, only services with >10 cases (minimal required number of observations for statistical analysis) were analyzed and included in the map: 95 cases responded by 37 service stations were removed. Fig. 3(1) shows that about half of the services have a coverage index of 0.75 or higher. The index values were randomly distributed across the state without obvious patterns. Fig. 3(2) shows that almost three-quarters of the services had a timeliness of 0.85 or higher, which means most of the stations responded within 15 min of the presumed 15-minute coverage areas. Stations with a high performance index are concentrated around Interstate highways I-29 and I-90.

Both performance indices represent the service-based performance from two perspectives: service coverage ratio and timely service rate. Travel distance is the main factor impacting response time; therefore, optimizing EMS service location to improve service coverage will ultimately improve service performance. It is possible that factors other than distance contribute to the variation of response time. A regression analysis can help identify factors significantly impacting response time (e.g., weather, roadway conditions) and can also identify ways of improving timely services. The following sections discuss the related regression methods and analysis.

## 3. Data preparation for regression models

Each 911 record had at least 50 case-specific attributes (i.e., attributes associated only with each 911 case) or service-specific attributes (i.e., attributes associated only with each EMS stations). After a careful review of all available attributes, the following were considered as case-specific variables: *caller's complaint*, *light and siren*, *dispatch time*, *location type*, *and weather*. Response time can be affected by many factors, such as the severity of the incident, EMS station location and staffing, weather, highway, and traffic conditions.

The "caller's complaint" can significantly impact response time, because urgent, life-threatening incidents such as strokes, breathing

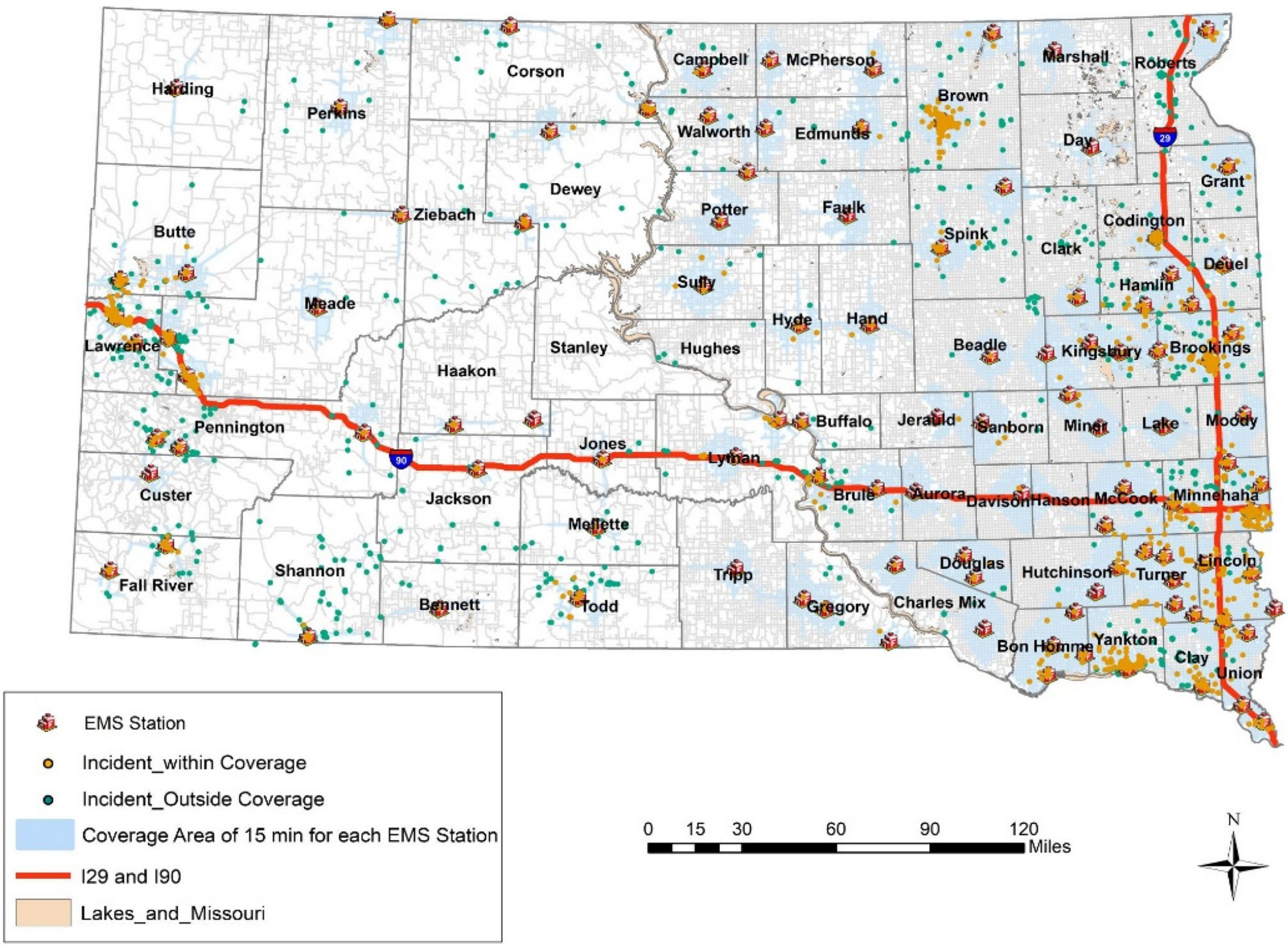



**Fig. 1.** EMS coverage map.

problems, and cardiac arrests demand shorter response time. The use of ambulance "light and siren" indicates the 911 response is urgent or that the driver is trying to avoid traffic congestion. "Dispatch time" includes the time of day and day of the week for each dispatch and may be impacted by service demand, light conditions, and staffing status. "Location type" indicates whether the incident is in a public area. Compared with a private address, a public area such as a nursing home usually has better location information and more convenient access points for an ambulance. This study extracted "weather" data such as mean visibility, mean wind speed, and weather indicator (if there is rain, snow or fog) from Weather Underground (WU) (https://www.wunderground.com/). Each event adopted the weather condition from the nearest weather service station on the same day.

Only life-threatening or severe cases (i.e. strokes, breathing problems, and cardiac arrests) were included in the analysis because for these cases, 1) the response time is more sensitive to a patient's outcome; and 2) event address information is more accurate. The response time of 8 min is considered a well-accepted criterion across United States EMS regions, especially for those life-threatening cases [28]. Thus, severe cases within each station's 8-minute coverage area were analyzed to explore the factors contributing to long travel time. Eight hundred and eighty-one (881) severe cases with 63 responding EMS stations met the inclusion criteria.

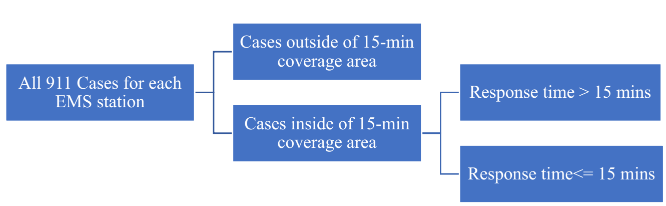


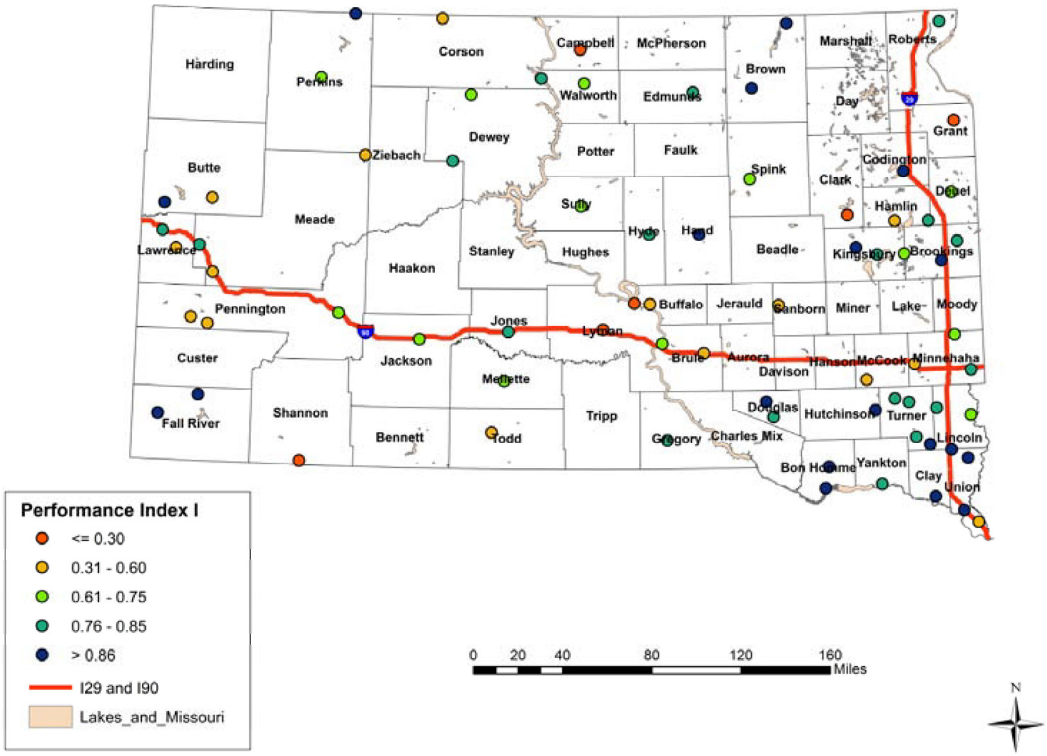

**Fig. 2.** Illustration of two performance indexes.

Several EMS service-specific variables were also considered in the regression analysis, including whether an EMS service is staffed with professional emergency medical technicians, vehicles and medical equipment at the station, and the proximity of streets and highways. Relevant literature suggests road density and connectivity (e.g., the number of nodes divided by the number of links) are accessibility indicators, while average traveling speed is a mobility indicator [29]. From this stance the authors based the roadway accessibility and mobility indexes on a highway network created by an 8-minute travel distance from the associated EMS station. The accessibility and mobility indicators for each station are formulated from Eqs. (1) to (3) as follows [29]:

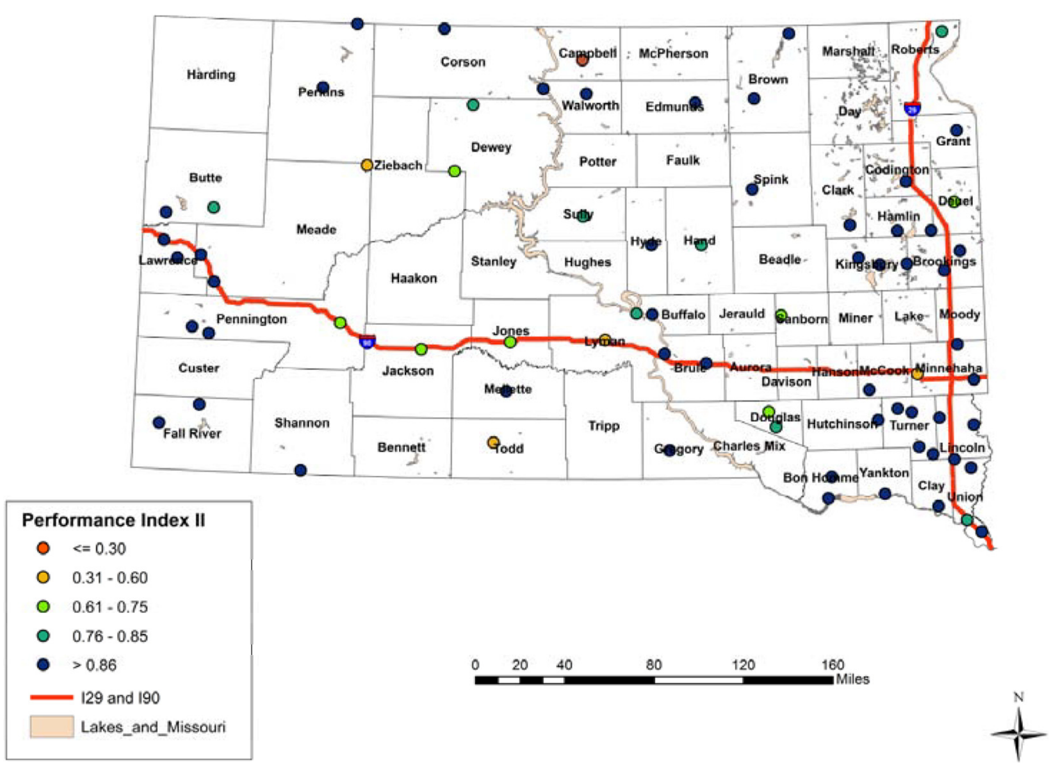
$$Density = \frac{\sum link\ length}{covered\ area} \tag{1}$$

$$Connectivity = \frac{number\ of\ links}{number\ of\ nodes} \tag{2}$$

$$Speed = \frac{\sum(link\ speed \times link\ length)}{\sum link\ length} \tag{3}$$

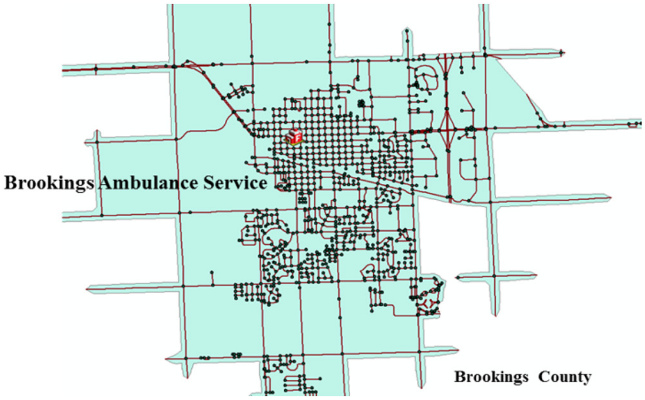



1) Service Coverage Index




2) Service Timeliness Index

**Fig. 3.** Performance indexes for each EMS station.




| | |
|---|---|
| Number of links | 1,467 |
| Number of nodes | 1,079 |
| Total link length (mile) | 94.39 |
| Total area (mile$^2$) | 25.87 |
| Average speed (mph) | 32.33 |
| Highway density (mile$^{-1}$) | 3.65 |
| Highway connectivity | 1.36 |

**Fig. 4.** Accessibility and mobility of the EMS station in Brookings County.

The roadway length, number of links, and number of nodes were derived from the South Dakota state highway links and nodes map in ArcGIS by clipping an 8-minute polygon for each EMS station. Fig. 4 uses Brookings County as an example to show the clipped area for an EMS station along with the variables used to calculate highway mobility and connectivity.

Besides highway accessibility and mobility, workload factor is included as another service-specific attribute to show level of EMS activity. Two types of service workload were introduced: yearly call volume and unit hour utilization (UHU) [30]. Yearly call volume is simply the number of 911 calls received by an EMS station. UHU is the length of time one ambulance unit is occupied over the total amount of time (i.e., 24/7 and 365 days a year). UHU is formulated in Eq. (4) [30]:

$$UHU = \frac{D \times Total\ Time}{n \times 8760} \tag{4}$$

where,

D is the yearly demand for each station,

TotalTime is the average total time in hour for each station and,

n is the number of ambulances for each station.

As a common practice, EMS services with a UHU below 35% are considered less active [30]. UHU was calculated for each EMS service, with results ranging from 2% to 10%. These results are expected because these are less-populated areas in the rural countryside.

Both case-specific and service-specific variables were calculated. Table 1 lists the variables and their descriptions.

## 4. Methods

To capture the relationship between response time and its contributing factors (including both case-specific attributes and station-specific attributes) for the 881 severe cases, three types of regression models: linear regression, spatial econometric models, and Geographically Weighted Regression (GWR) were used. The latter two are capable of handling spatial autocorrelation in the data. To choose the best model fitting South Dakota EMS data, the regression models were evaluated by measure of spatial autocorrelation and measure of goodness-of-fit. Instead of using the aggregated values of time over a predefined area [18], this study focuses on response time related to individual incidents using these statistics, which can accurately reveal the underlying factors affecting response time for each event. The rest of this section discusses the regression models, model assessment and comparison.

### 4.1. Linear regression

The linear regression model can generate a linear relationship between a dependent variable and multiple independent variables. The model assumes that the residuals are independent, normally distributed, and have a mean of zero and constant variance. The equation is shown in Eq. (5):

$$y = \beta_0 + \sum_{k=1}^{p} \beta_k x_k + \varepsilon \tag{5}$$

**Table 1**
Variable description.

| Variable | | | Description | NHTSA 2 code[a] |
|---|---|---|---|---|
| Dependent variable | | ERTime | Response time (minutes) | |
| Independent variable | Case-specific | Caller's complaint | Severe (1) or not (0) | E03_01 |
| | | Response mode | Light/siren on (1) or not (0) | E02_20 |
| | | Time of day | Day (1) or night (0) | E05_03 |
| | | Day of week | Weekday (1) or weekend (0) | E05_03 |
| | | Location | Public area (1) or not (0) | E08_07 |
| | | Visibility | Mean visibility (miles) | |
| | | Wind speed | Mean wind speed (mph) | |
| | | Weather indicator | Extreme weather such as rain, snow (1) or not (0) | |
| | Service-specific | Density | Highway density | |
| | | Connectivity | Highway connectivity | |
| | | Speed | Average speed or posted speed limit (mph) | |
| | | Professional | Professional (1) or volunteer (0) | |
| | | Vehicle | Number of ambulance | |
| | | EMS demand | 911 call volume | |
| | | UHU | Unit hour unitization | |

[a] NHTSA 2 Code for items in NEMSIS EMS dataset to retrieve information for case-specific variables (https://doh.sd.gov/documents/EMS/DataDictionary.pdf).

where $y$, $x_k$, ε indicate dependent variable, kth independent variable and the normal error respectively; and coefficients $\beta_k$ are the global parameters.

*4.2. Spatial econometric models*

Spatial econometric models consider the spatial autocorrelation effect from neighbors, which requires a spatial weight matrix of neighbor data points. The matrix is generated using neighbors that share the same border, or the k-nearest neighbors, or neighbors within a certain distance. Two popular models that address spatial autocorrelation in response variable and residual respectively are the spatial lag model and the spatial error model (shown in Eqs. (6) and (7)).

$$\text{Spatial Lag Model}: Y = \rho WY + X\beta + \varepsilon \tag{6}$$

$$\text{Spatial Error Model}: Y = \rho W(Y - X\beta) + X\beta + \varepsilon \tag{7}$$

where, Y, X, and ε indicate response variable, independent variables and identical independent error term respectively. $\beta$ is the vector of coefficients and W is the spatial weight matrix. $\rho$ denotes the autoregressive parameter.

*4.3. Geographically weighted regression (GWR)*

Compared to a linear regression model, the GWR model considers that parameters vary across space. The purpose of GWR is to estimate different relationships between the dependent and independent variables for each geographic location. Local parameters are used in the GWR model, which is formulated in Eq. (8):

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{k,i} + \varepsilon_i \tag{8}$$

where $y_i$, $x_{k,\ i}$, $\varepsilon_i$ indicate dependent variable, kth independent variable and the normal error at location i respectively; $(u_i, v_i)$ is the coordinate of the ith location; and coefficients $\beta_k(u_i, v_i)$ are the local parameters at the location [19]. Based on the concept of GWR models, the local parameters $\beta_k(u_i, v_i)$ ($k$: 0, 1, …, $p$) are estimated for each location i, thus $n * (P + 1)$ parameters are estimated for $n$ observations.

Apart from traditional GWR, the extension of GWR (mixed GWR) can have both global variables and local variables in the model structure. In GWR modeling, the local parameters for each location can be estimated based on observations from nearby locations. A location's parameters are more strongly affected by the observations occurring close by as opposed to observations made from farther away. The influence factor is called the weighting function $w_{ij}$. Two commonly used weighting functions, Gaussian and bi-square, are listed below:

$$\text{Gaussian}: w_{ij} = e^{-\frac{d_{ij}^2}{h^2}} \tag{9}$$

$$\text{Bi-square}: w_{ij} = \begin{cases} \left(1 - \left(\dfrac{d_{ij}}{h_i}\right)^2\right)^2; \ if\ d_{ij} < h_i \\ 0;\ otherwise \end{cases} \tag{10}$$

where $d_{ij}$ is the distance from location i to location j, h and $h_i$ are the bandwidth for these two functions [19].

Bandwidth for the Gaussian function is constant, meaning the magnitude of the function is the same for each location [19]. Bandwidth for the Bi-square function $h_i$ varies across locations and is defined as the nth nearest observation from location I [19]. Bi-square's adaptive function is often used when data are not distributed randomly because this function can be adjusted to the density of data. The Akaike information criterion (AIC) is often used to select the optimal bandwidth and the best model; models with a lower AIC perform better [19].

*4.4. Model assessment & comparison*

*4.4.1. Measure of spatial autocorrelation*

Moran's I, is a spatial statistical method that uses feature location and feature values to measure whether spatial autocorrelation exists across the entire study area. The value of Moran's I index denotes whether it shows clustered, dispersed, or randomly distributed patterns in the study area [31]. *Z*-score and p-value are generated to evaluate the index significance. A large enough absolute value of z-score or low enough p-value indicates a failure to accept the null hypothesis, meaning a clustered or dispersed pattern exists; a low enough absolute value of z-score means the null hypothesis cannot be rejected, and that a randomly distributed pattern exists.

*4.4.2. Measures of goodness of fit*

Several goodness of fit measures can assess model performance: R-squared, Akaike information criterion (AIC), Mean absolute deviation (MAD), and Mean squared prediction error (MSPE). R-squared, the most popular measure, indicates the percentage of variation in the dependent variable which can be explained by the independent variables. The higher the R-square, the better the model performance. A lower AIC indicates that the predicted values are closer to the real values. MAD and MSPE measure the accuracy of the predicted values, with a lower value suggesting better accuracy.

## 5. Findings & discussion

*5.1. Linear regression*

The linear regression analysis began with seven case-specific variables and seven station-specific variables. A collinearity analysis was conducted for all independent variables, and the results suggested that correlations exist in the following three pairs: *Professional* and *Vehicle* (Pearson correlation coefficient = 0.83), *Vehicle* and *Demand* (Pearson correlation coefficient = 0.78), and *Demand* and *UHU* (Pearson correlation coefficient = 0.74). Stepwise selection was used to keep the statistically significant variables in the final model. The final model includes six variables and is specified as: *Response time ~ Response Mode + Location + Mean Visibility + Highway Connectivity + Professional + EMS Demand*. Descriptive statistics of the variables are shown in Table 2. Continuous variables (*Mean Visibility*, *Highway Connectivity and EMS Demand*) have the mean value, standard deviation, minimum value and maximum value, and binary variables (*Response Mode*, *Location and Professional*) have the percentage of category “1”.

The R software was used to develop this linear model. Parameter estimates and model performance are shown in Table 2. The positive sign for *Response Mode* means that response time increased when the light and siren were on, suggesting possible traffic congestion. Incidents happening in public areas had a reduced response time. *Mean Visibility* had a negative effect on the response time. Response time decreases with the increase of road accessibility (*Highway Connectivity*). EMS services with professional staff show a shorter response time compared to stations using *only* volunteers. Higher *EMS demand*, which indicates that the service is busy, increased the response time. The Moran's I statistic (4.67) shows a positive statistically significant (p-value <0.001) spatial autocorrelation.

*5.2. Spatial econometric models*

A spatial lag model and a spatial error model were implemented in the R software to account for the spatial autocorrelation. The model began by including the six statistically significant variables obtained from linear regression and then added other variables one by one. Each time when a new variable was added, the variable significance and model AIC value were checked and compared to the previous model. After exhausting all the available variables, the performance of

**Table 2**
Comparison among models.

| Variables | | Descriptive statistics | Linear | | Spatial lag | | Spatial error | | GWR | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (Mean, STD, min, max) or % of "1" | Estimate | *t*-Test | Estimate | z-Test | Estimate | z-Test | Min | Median | Max |
| | | | 95% confidence interval | | 95% confidence interval | | 95% confidence interval | | | | |
| Dependent | Response time | (4, 4.1929, 1, 46) | | | | | – | | | | |
| | Intercept | – | 11.4641<br>(7.9874, 14.9429) | 6.4735 | 16.8194<br>(13.3465, 20.2893) | 5.3509 | 11.2519<br>(7.7778, 14.7313) | 6.8989 | −14.0532 | 7.9212 | 26.8969 |
| | Response mode | 24.97% of "1" | 0.8906<br>(0.2699, 1.5104) | 2.8201 | 0.8523<br>(0.2288, 1.4813) | 2.7196 | 0.8473<br>(0.2248, 1.4660) | 2.6406 | (Est., *t*-test) = (−0.335, 0.861) | | |
| | Location | 30.87% of "1" | −0.5529<br>(−1.1188, 0.0113) | −1.922 | −0.6292<br>(−1.1941, −0.074) | −2.1896 | −0.5518<br>(−1.117, 0.0133) | −1.9213 | −3.3015 | −0.573 | 0.4617 |
| Independent | Mean visibility | (9, 1.9360, 1, 10) | −0.2905<br>(−0.4239, −0.1567) | −4.2711 | −0.2846<br>(−0.4182, −0.1427) | −4.2204 | −0.2899<br>(−0.4187, −0.1619) | −4.285 | −2.0345 | −0.1103 | −0.0008 |
| | Road connectivity | (1.3690, 0.1168, 1.02, 1.74) | −3.8647<br>(−6.122, −1.6087) | −3.3632 | −5.3865<br>(−7.6478, −3.1403) | −3.9657 | −3.7111<br>(−5.9618, −1.4562) | −3.6409 | −14.9228 | −0.5099 | 12.5605 |
| | Professional | 55.28% of "1" | −1.6856<br>(−2.4198, −0.9538) | −4.5169 | −1.8124<br>(−2.5376, −1.0905) | −4.8361 | −1.6558<br>(−2.387, −0.9162) | −4.4034 | −2.9936 | −0.5139 | 2.9376 |
| | EMS demand | (987, 955.8463, 13, 4785) | 0.0018<br>(0.0014, 0.0022) | 9.2728 | 0.0018<br>(0.0014, 0.0023) | 9.2862 | 0.0018<br>(0.0014, 0.002) | 9.1573 | (Est., *t*-test) = (0.0013, 4.666) | | |

| Performance | Estimate | Estimate | Estimate | Estimate |
|---|---|---|---|---|
| R Square | 0.15 | 0.16 | 0.15 | 0.31 |
| MAD | 2.41 | 2.39 | 2.41 | 2.23 |
| MSPE | 15 | 14.84 | 15 | 12.09 |
| AIC | 4902 | 4846 | 4865 | 4773 |
| Moran's I | 4.67 | 3.86 | 4.67 | −2.22 |

the model with six original variables has not been improved by augmenting any new variable. Therefore, the same variables were used for spatial econometric models. Because 911 events were unevenly distributed across the state, the k- nearest neighbor method was chosen to calculate the weight matrix in which 251 of the nearest neighbor points were selected. Both autoregressive parameters $\rho$ were shown to be significant, suggesting the presence of spatial correlation. The signs of coefficient estimates in Table 2 are consistent with the linear model.

### 5.3. Geographically weighted regression (GWR)

The GWR model was formulated using GWR 4.0, a statistical software package specially developed for GWR [32]. The same variables were adopted because no new variables significantly improved model performance. Since the choice of weight function is vital, both Gaussian (Eq. (9)) and bi-square functions (Eq. (10)) were evaluated. GWR with the bi-square function outperformed the Gaussian function with lower AIC and higher R-squared values. A geographical variability test was performed for local coefficients to test whether they varied across the space. The results suggested several possible global variables, and thus, a mixed GWR model was applied.

In GWR 4.0, mixed GWR was initially set as a local model in which all the independent variables were treated as local variables. After an iterative process, some variables became global terms, and others remained local. An iterative golden section search of the AIC function revealed that when the number of nearest observations was 251, the AIC score was optimal; thus, the bandwidth was set as the 251 nearest observations. Because the local coefficient estimate varies across space, it is described by the range of value (minimum, median, and maximum) rather than the mean. By comparing lower quartile and upper quartile in Table 2, most signs of the parameters vary from negative to positive except for *Mean Visibility*. The global variables, *Response Mode* and *EMS Demand*, have a coefficient estimate similar to the linear regression estimate, but *Response Mode* is not statistically significant in the GWR model.

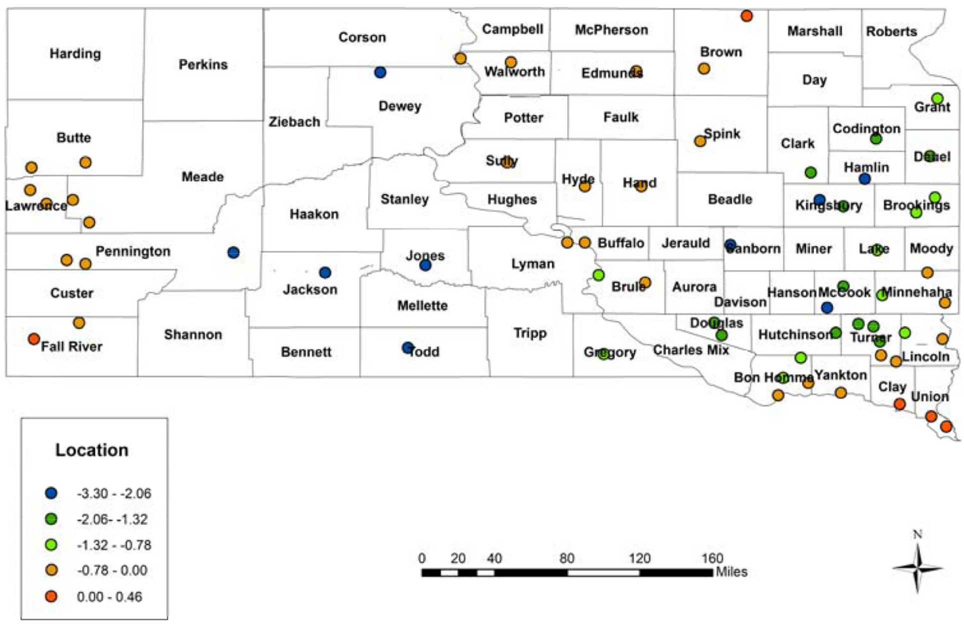
In GWR, maps can help researchers visualize the coefficient and the statistical significance measured by the t-statistics for each EMS service. The coefficient may vary in its sign, magnitude, and statistical significance, all of which should be taken into consideration for further analysis. Clusters on the Fig. 5 map indicate similar coefficient estimates. For the *Location* variable, few services located on the southeast side of the state have a positive coefficient estimate on response time; however, the variable at these stations is not statistically significant. In other words, *Location* has a negative effect on response time when it is statistically significant. Although the effect is negative across the entire state for the *Mean Visibility*, the Southwest and South-Central regions seem to have a greater impact on response time. Some stations on the east or west of the state show a positive effect for *Highway Connectivity*, but not all effects are statistically significant. *Professional* seems to be associated with reduced response time except for stations in the west region. A look at the t value shows that the coefficients for western stations are not statistically significant.

### 5.4. Discussion

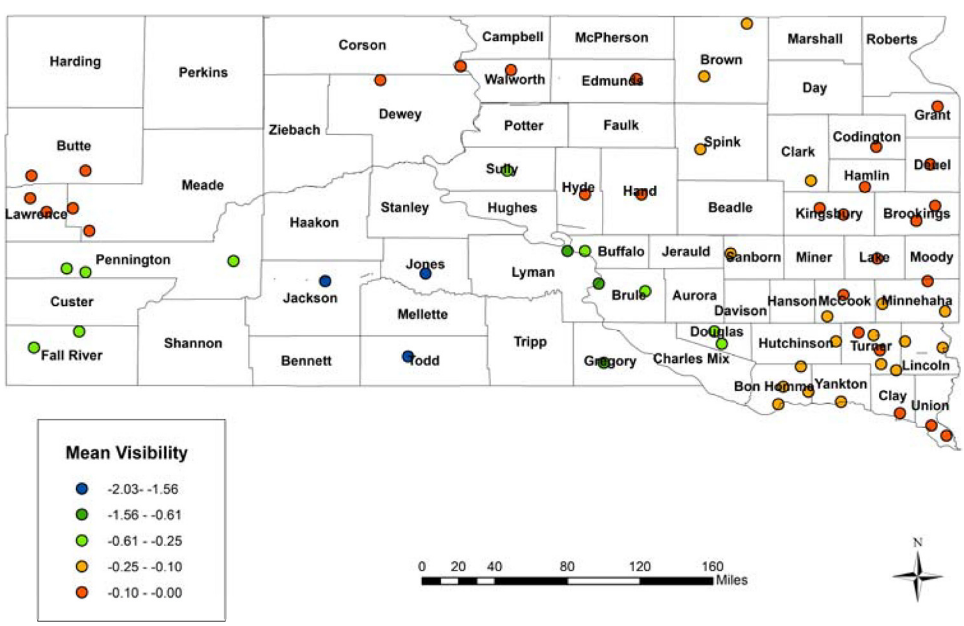
Both the spatial econometrics model and GWR model can improve model performance. The GWR model improved R-squared from 0.15 to 0.31, while spatial lag or spatial error shows marginal improvement. GWR ranks first in prediction accuracy with the lowest MAD, MSPE and AIC, followed by the spatial lag model and the spatial error model. The decreased Moran's I after applying spatial models indicates that both the GWR and spatial econometrics models can capture spatial autocorrelation.

*Response Mode* has a positive impact in both the linear and spatial econometric models, meaning increased response time is associated with light and siren situations. The factor is no longer statistically significant in GWR. Note that GWR is a local regression analysis in which the parameters for each location are calibrated from observations from nearby locations, not all the observations. At the local level, the choice of response mode is not significantly correlated with the response time but, at the global level, a positive impact is observed. All models show the same sign for *Mean Visibility* and *EMS Demand*. *Mean Visibility* has a negative impact on response time, while *EMS Demand* increases the response time. *Location* has a negative impact on response time for stations where the variable is statistically significant in GWR; this result is consistent with linear regression results, suggesting that incidents happening in public areas are associated with shorter response times. A negative effect is observed in the linear regression and spatial econometric models concerning *Highway Connectivity*; however, *Highway Connectivity* in GWR is positive for some stations and negative for others. Similarly, the *Professional* coefficient is negative for the linear regression and spatial econometric models but it varies in GWR. The disparities among stations in terms of coefficient estimates underscore the importance of using GWR as an analytical tool to investigate further of the nature of local variations in relationships.

The strength of GWR is its ability to offer geointelligence on the effects and statistical significance of each variable. Variables may have statistically significant relationship at some services but not at others. Therefore, GWR is helpful in exploring the following questions: Is the relationship intrinsically different across space; perhaps, there are spatial variations in training, service protocols, other drivers' response to emergency vehicles and highway traffic conditions? Is the spatial non-stationarity caused by the omission of some important variables (e.g., local agencies' operational practices), which can prompt further

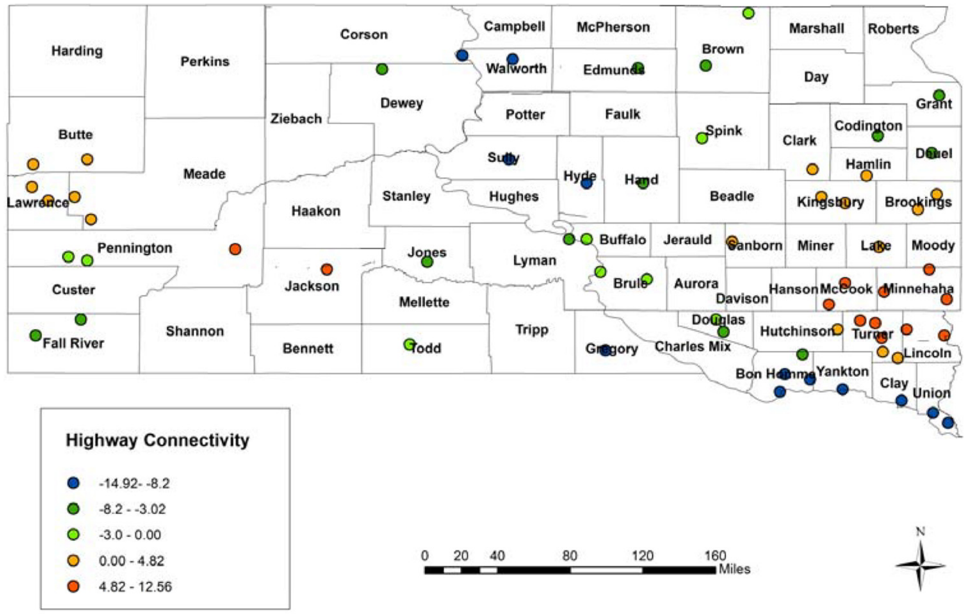



(1) Location

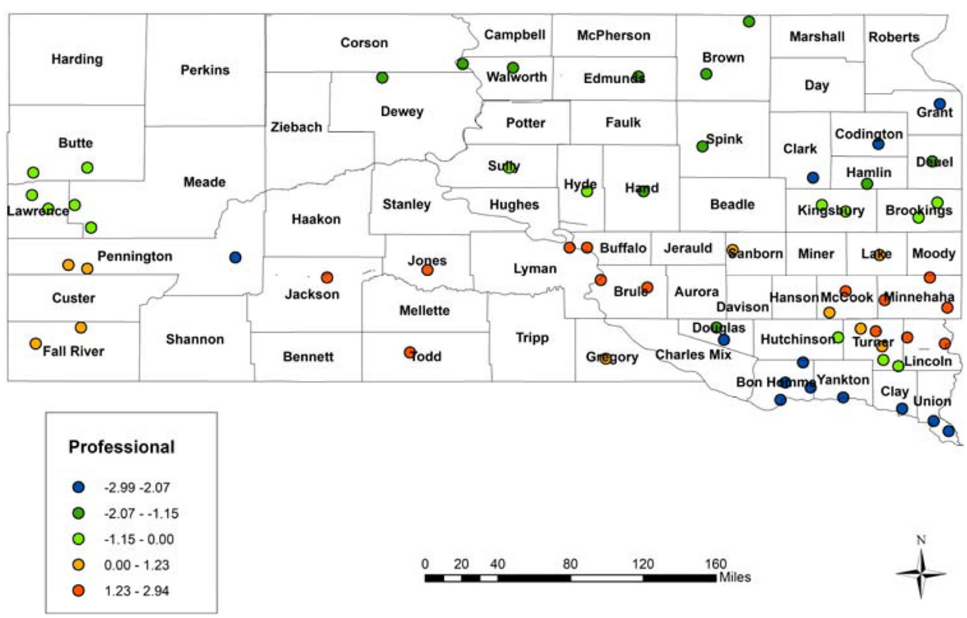



(2) Mean Visibility

**Fig. 5.** GWR coefficient map for local variables.




(3) Highway Connectivity




(4) Professional

**Fig. 5** (*continued*).

investigation and inspire new research? Table 3 shows each selected station and its corresponding significant variables that affect service performance. For example, Brookings ambulance service has two significant variables (*Location* and *EMS Demand*) which can explain only <10% of the variation in response time, suggesting further investigation of unobserved factors is necessary. The revelation from the spatial models suggests a need to review more detailed information before attempting to reach a definitive conclusion, especially for stations with subpar performance.

## 6. Conclusions

This study established two data-driven performance metrics for EMS: the service coverage ratio index and the service timeliness index using NEMSIS data and identified statistically significant factors contributing to EMS response time from a list of fifteen (15) input variables, including eight (8) case-specific and seven (7) station-specific variables. Given the large amount of information available through NEMSIS, this study shows one method, GWR, is useful to disaggregate the data and

**Table 3**
Significant contributing factors affecting response time for selected stations.

| EMS station | Response mode | Location | Mean visibility | Highway connectivity | Professional | EMS demand | Local$R^2$ |
|---|---|---|---|---|---|---|---|
| Watertown fire dept & ambulance service | – | −1.5124 | – | – | −2.5433 | 0.0013 | 0.0877 |
| Aberdeen fire & rescue | – | – | – | – | −1.5983 | 0.0013 | 0.0465 |
| Lead-deadwood regional hospital ambulance | – | – | – | – | – | 0.0013 | 0.1181 |
| Dell rapids community ambulance service | – | – | – | – | – | 0.0013 | 0.0900 |
| Yankton county EMS | – | – | – | −8.4235 | −2.5568 | 0.0013 | 0.0666 |
| Brookings ambulance service | – | −1.1487 | – | – | – | 0.0013 | 0.0921 |
| Clark county ambulance service | – | −1.6721 | – | – | −2.5186 | 0.0013 | 0.0875 |
| Hand county ambulance service | – | – | – | −6.9439 | – | 0.0013 | 0.0501 |
| Hand county ambulance service | – | – | – | −8.9216 | – | 0.0013 | 0.0427 |
| Dell rapids community ambulance service | – | −2.5354 | −2.0345 | 7.4267 | – | 0.0013 | 0.5141 |
| Springfield fire & ambulance service | – | – | – | −9.5200 | −2.9936 | 0.0013 | 0.0945 |
| Madison medical services | – | −1.1839 | – | – | – | 0.0013 | 0.0924 |

focus on response time variables. This in turn should help state, regional, and local EMS leadership with strategic planning including: service locations/re-locations and subsequent logistical/structural elements involved in quality improvement efforts.

In this study, response time was used as the performance metric and a 15-minute response time was adopted as the benchmark. Service coverage and timely service performance indexes were developed to evaluate the positioning and service quality of each EMS station. Low service coverage measure suggests the need for more strategic establishment or relocation of service stations. Low timely service measure suggests EMS decision makers should identify other factors besides the distance that impacts the response time. Results show a well-positioned station with well-trained staff should be able to respond to more 911 calls within the 15-min time frame and should also have a higher percentage of successes if a 911 call is located within the estimated 15-min coverage area.

Several regression models were developed and compared for response time: the linear regression model, the spatial lag model, the spatial error model, and the GWR model. Statistically, the GWR model performed better than the linear regression and spatial econometric models (spatial lag and spatial error). Geographically, GWR has the ability to explore the underlying variations in relationships of a set of variables over space, whereas this information cannot be disclosed by global models such as the linear regression model. The GWR results provide additional insights into the location-specific spatial patterns of coefficient estimates and their statistical significance. Local practitioners may focus on different aspects and identify new measures to reduce response time. The model using GWR not only offers state EMS officials an overview of the statistically significant factors affecting response time across the state, but also provides a good reference for local agencies seeking solutions for shortening response time at the station level.

Finally, the tremendous value of the EMS data through information extraction, knowledge discovery and performance metrics development underscores the importance of improving data quality. Inaccurate data means a waste of time and resource. More importantly, an improved dataset will provide more reliable and relevant information that guide effective EMS planning and operations.

## References

[1] Gamm L, et al. Rural healthy people 2010: identifying rural health priorities and models for practice. J Rural Health 2002;18(1):9–14.
[2] NHTSA. Traffic safety facts 2010: a compilation of motor vehicle crash data from the Fatality Analysis Reporting System and the General Estimates System. DOT HS 811 659; 2012.
[3] FHWA. Highway Statistics. Highway statistics series 2014. [cited 2015 February 10]; Available from http://www.fhwa.dot.gov/policyinformation/statistics/2010/; 2010.
[4] NHTSA. Promoting safe passage into the 21 Century. SDDOT; 1998.
[5] Becknell JM, Simon L. Beyond EMS data collection: envisioning an information-driven future for emergency medical services. National Highway Traffic Safety Administration; 2016.
[6] Solutions S. A Crisis and Crossroad in rural North Dakota emergency medical services; 2011, June.
[7] Renger R, et al. Challenges and solutions facing EMS in supporting the IOM recommendation for a national cardiac arrest registry: A system perspective. Center for Rural Health, School of Medicine & Health Sciences, University of North Dakota; 2016.
[8] Busko J, Rural EMS. Emergency medical services: clinical practice and systems oversight. Dubuque, IA: Kendall-Hunt; 2008; 218–28.
[9] Dawson DE. National emergency medical services information system (NEMSIS). Prehosp Emerg Care 2006;10(3):314–6.
[10] NHTSA. Emergency medical services performance measures: recommended attributes and indicators for system and service performance; 2009.
[11] Prevention CfDCa. Motor vehicle safety. Injury and violence prevention; 2013.
[12] Al-Shaqsi SZK. Response time as a sole performance indicator in EMS: pitfalls and solutions. Open Access Emerg Med 2010;2:1–6.
[13] Anselin L. Spatial econometrics: methods and models. , vol. 4Springer Science & Business Media; 2013.
[14] Efthymiou D, Antoniou C. Investigating the impact of recession on transportation cost capitalization: a spatial analysis. J Transp Geogr 2015;42:1–9.
[15] Wang C-H, Chen N. A GIS-based spatial statistical approach to modeling job accessibility by transportation mode: case study of Columbus, Ohio. J Transp Geogr 2015; 45:1–11.
[16] Ribeiro A, Antunes AP, Páez A. Road accessibility and cohesion in lagging regions: empirical evidence from Portugal based on spatial econometric models. J Transp Geogr 2010;18(1):125–32.
[17] Wang Y, Kockelman KM. A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accid Anal Prev 2013;60:71–84.
[18] Dimitriou L, Efthymiou D, Antoniou C. Understanding the impact of accessibility and weather on emergency unit reaction times. Transportation research board 95th annual meeting; 2016.
[19] Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted regression. John Wiley & Sons, Limited; 2003.
[20] Zhao F, Park N. Using geographically weighted regression models to estimate annual average daily traffic. Transp Res Rec 2004;1879:99–107.
[21] Du H, Mulley C. Relationship between transport accessibility and land value: local model approach with geographically weighted regression. Transp Res Rec 2006; 1977:197–205.
[22] Li Z, et al. Using geographically weighted Poisson regression for county-level crash modeling in California. Saf Sci 2013;58:89–97.
[23] Selby B, Kockelman KM. Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression. J Transp Geogr 2013;29:24–32.
[24] Pirdavani A, et al. Assessing the road safety impacts of a teleworking policy by means of geographically weighted regression method. J Transp Geogr 2014;39:96–110.
[25] Qin X, He Z, Samra H. Rural emergency medical service needs assessment. Transportation research board 94th annual meeting; 2015.
[26] LLP", S.S. An introduction to rural EMS medical direction in South Dakota. The South Dakota Office of Rural Health; 2011.
[27] ArcMap, A. 10.2. Environmental systems research institute, Redlands, CA; 2013.
[28] Association, N.F.P., NFPA 1500. Standard on fire department occupational safety and health program. 1992: National Fire Protection Association; 2007.
[29] Litman T. Evaluating accessibility for transportation planning; 2008.
[30] TriData. Assessment of fire department resource locations, staffing, and facilities. Division of System Planning Corporation; 2007.
[31] Ripley BD. Spatial statistics. , vol. 575John Wiley & Sons; 2005.
[32] Nakaya T. GWR4 user manual. GWR 4 Development Team; 2014.