

# Predictor Variables Influencing Visibility Prediction Based on Elevation and Its Range for Improving Traffic Operations and Safety

Ajinkya Sadashiv Mane<sup>1</sup>, Srinivas Subrahmanyam Pulugurtha<sup>1</sup>, Venkata Ramana Duddu<sup>1</sup>, Christopher Michael Godfrey<sup>2</sup>

<sup>1</sup>IDEAS Center, Civil & Environmental Engineering Department, The University of North Carolina at Charlotte, Charlotte, USA

<sup>2</sup>Atmospheric Sciences Department, The University of North Carolina at Asheville, Asheville, USA

Email: ajinkyamane@gmail.com, sspulugurtha@uncc.edu, vduddu@uncc.edu, cgodfrey@unca.edu

**How to cite this paper:** Mane, A.S., Pulugurtha, S.S., Duddu, V.R. and Godfrey, C.M. (2022) Predictor Variables Influencing Visibility Prediction Based on Elevation and Its Range for Improving Traffic Operations and Safety. *Journal of Transportation Technologies*, 12, 439-452.

<https://doi.org/10.4236/jtts.2022.123027>

**Received:** May 31, 2022

**Accepted:** July 18, 2022

**Published:** July 21, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Low visibility condition hinders both air traffic and road traffic operations. Accurate forecasting of visibility condition helps aircraft operators and travelers to make better decisions and improve their safety. It is, therefore, essential to investigate and identify the predictor variables that could influence and help predict visibility. The objective of this study is to identify the predictor variables that influence visibility. Four years of surface weather observations, from January 2011 to December 2014, were collected from the weather stations located in and around the state of North Carolina, USA for the model development. Ordinary least squares (OLS) and weighted least squares (WLS) regression models were developed for different visibility and elevation ranges. The results indicate that elevation, cloud cover, and precipitation are negatively associated with the visibility in visibility less than 15,000 m model. The elevation, cloud cover and the presence of water bodies within the vicinity play an important role in the visibility less than 2000 m model. The chances of low visibility condition are higher between six to twelve hours after the rainfall when compared to the first six hours after the rainfall. The results from this study help to understand the influence of predictor variables that should be dealt with to improve the traffic operations and safety concerning the visibility near the airports/road transportation network.

## Keywords

Visibility, Prediction, Weather Station, Linear Regression Model, Elevation

## 1. Introduction

Inclement weather conditions such as snow, sleet, fog, heavy rainfall, and cross-

wind affect the air and road traffic operations. From the year 1982 through the year 2013, inclement weather condition contributed to 35% of the general fatal aviation incidents in the United States [1]. Low ceiling, fog, and clouds were among the top three variables which contributed to the weather-related fatalities in air traffic operations. According to the Federal Highway Administration (FHWA), from the year 2005 to the year 2014, around 28,533 road crashes occurred per year due to fog, resulting in over 495 fatalities and 10,448 injury crashes [2]. From the year 2002 to the year 2012, there were 19,188 reported fog-related crashes in North Carolina alone [3]. A recent study also quantified the effect of rainfall and visibility conditions on road traffic travel time reliability [4]. Therefore, it is necessary to alert the aircraft operators and motorists beforehand to improve safety and mobility during low visibility. However, fog is a localized phenomenon. It would be expensive to install visibility sensors at every few miles along roads. Hence, identifying the predictor variables associated with low visibility not only helps predict visibility but also helps disseminate potential risk associated with low visibility condition.

In the past, researchers used several statistical and numerical modeling techniques for fog assessment. Vislocky and Fritsch [5] developed linear regression models to predict ceiling height and visibility. The threshold values for visibility were less than 1.61 km, 4.83 km, 8.05 km, and 11.27 km, which were based on aircraft operations. Surface observation parameters such as opaque cloud amount, cloud cover, precipitation occurrence, wind direction and speed, sea level pressure, dewpoint, and dewpoint depression were considered as the predictor variables. Hilliker and Fritsch [6] developed models to forecast visibility at the San Francisco International Airport for 1 - 6 hour lead times. They observed that the inclusion of upper-air predictor variables can reduce prediction error by 3% than models solely from surface data.

A detailed literature review related to the fog prediction methods is presented by Gultepe *et al.* [7]. Several research studies documented the parameters in fog assessments. Meyer *et al.* [8] showed that visibility in foggy conditions is a function of droplet number concentration. On the other hand, Jiusto [9] suggested that visibility is a function of both droplet size and liquid water content, concluding that liquid water content is directly related to the droplet size.

Tardif and Rasmussen [10] analyzed meteorological factors and scenarios leading to the occurrence of precipitation fog in the New York City area. Their study indicates that 18% of the analyzed precipitation events corresponded with fog events and that the majority of fog events occurred during light precipitation. Most of the fog events occurred at high elevation stations due to upslope flow and lowering of the cloud base. Since relative humidity is a function of temperature, they divided fog events into those that occurred due to moistening, cooling, moistening and cooling, or static conditions. An analysis of all fog events based on these tendencies indicates that moistening, cooling, moistening and cooling, and static conditions were observed for 42%, 25%, 10%, and 23% of the fog

events, respectively.

Studies such as by Pulugurtha *et al.* [11] have explored the influence of surface weather observations and meteorological predictor variables on visibility. However, the visibility could be influenced by time-of-the-day, rainfall in past hours, wind speed, and the presence of water bodies within the vicinity. The contribution of the aforementioned predictor variables for different visibility and elevation ranges could vary and has not been explored in the past. Therefore, this study focuses on identifying the meteorological and temporal predictor variables which could influence the visibility with respect to change in the elevation and its range.

## 2. Methodology

The methodology adopted in this study includes the following steps.

- 1) Identify the weather stations and collect surface weather observations
- 2) Process data
- 3) Develop and compare the linear regression models
- 4) Validate the developed models

Each step is explained next in detail.

### 2.1. Identify the Weather Stations and Collect Surface Weather Observations

National Oceanic and Atmospheric Administration (NOAA)/National Centers for Environmental Information (NCEI) collects hourly meteorological data from over 20,000 locations across the world [12] [13]. This Integrated Surface Database (ISD) includes visibility, 2-m air temperature, dew point temperature, wind speed, atmospheric pressure, precipitation, and current weather conditions. Some stations also collect snow depth and snowfall information. The ISD database undergoes a meticulous quality control process before distribution [14] [15]. However, data quality issues still remain in the database [15]. These issues are dealt with effectively through additional quality assurance algorithms.

Data for four years, from January 2011 to December 2014, for 238 ISD locations in and near the state of North Carolina, USA were collected and processed for this study.

### 2.2. Process Data

The collected surface weather observations were processed by deleting the missing values and outliers using Microsoft SQL server. Further, precipitation in previous hours and time-of-the-day could influence the formation of fog. Therefore, binary variables such as the occurrence of rainfall and time-of-the-day were added to the database.

Oliver [3] stated that crashes related to the low visibility conditions (49%) are more likely during morning hours. In addition, the literature review indicated that precipitation is a governing factor in low visibility conditions. Therefore,

the occurrence of rainfall variable was classified into four groups. They are rainfall in the past 3 hours (R0-3hrs), 3 to 6 hours (R3-6hrs), 6 to 12 hours (R6-12hrs), and 12 to 24 hours (R12-24hrs). Also, six subcategories of time-of-the-day were identified. They are 12 AM to 4 AM (T0am-4am), 4 AM to 8 AM (T4am-8am), 8 AM to 12 PM (T8am-12pm), 12 PM to 4 PM (T12pm-4pm), 4 PM to 8 PM (T4pm-8pm), and 8 PM to 12 AM (T8pm-12am). Each of the categories was considered as a binary variable. In addition, dewpoint depression was computed by taking the difference between the air temperature and dew point temperature. Typically, fog forms when the dewpoint depression is roughly less than 2.5°C - 4.0°C.

According to the NOAA, fog is formed by the collection of suspended water droplets or ice crystals near Earth's surface, which reduces the horizontal visibility below 1 km [16]. The water bodies in the vicinity of the weather station could influence the formation of fog. Therefore, the presence of water bodies within a 1.61-km buffer of each weather station was captured using ArcGIS and was represented as a dichotomous variable.

### 2.3. Develop and Compare the Regression Models

Ordinary least squares (OLS) and weighted least squares (WLS) regression models were developed to investigate the effect of predictor variables on the visibility. OLS regression is the simplest form of linear regression and tries to minimize the residual sum of squares (RSS). In OLS regression, equal weight is given for each observation to minimize the RSS. However, WLS regression minimizes the weighted RSS with weight<sub>*i*</sub> = 1/variance<sub>*i*</sub> [17], where “*i*” is the individual observation. In other words, more weights are given to the observations closer to the population mean.

OLS and WLS regression models were developed based on visibility and elevation. The visibility range was classified into four groups: less than 15,000 m, less than 10,000 m, less than 5000 m, and less than 2000 m. As per the literature review, change in the elevation alters the visibility. Therefore, regression models were developed by the elevation range. The elevation ranges were classified into five groups: less than 50 m, 50 m to 250 m, 250 m to 750 m, and greater than 750 m.

For each visibility range, regression models were developed by considering all the samples irrespective of the elevation. Overall, twenty OLS and twenty WLS regression models were developed to investigate the influence of predictor variables on visibility based on change in the elevation.

Predictor variables with a level of significance (p-value) less than 0.05 (at a 95% confidence level) were considered to have a statistically significant effect on the dependent variable (visibility). Predictor variables with p-value more than 0.05 were removed from the regression model one at a time. This method is also called as a backward elimination method. Statistical measures such as R-squared, adjusted R-squared, Akaike Information Criterion (AIC), and root mean square

error (RMSE) were computed to evaluate the performance of the models developed.

## 2.4. Validate the Developed Models

Among all the developed models, the best-fitted model was selected based on the statistical tests. Two months (July and August of 2016) of hourly surface weather observations from the weather stations located in and around the state of North Carolina were collected for validation. Mean Absolute Percentage Error (MAPE) was used to check the accuracy of the developed model. The mathematical expression of MAPE is expressed as Equation (1).

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{V_{\text{observed}(i,j)} - V_{\text{predicted}(i,j)}}{V_{\text{observed}(i,j)}} \right| \quad (1)$$

where,  $V_{\text{observed}(i,j)}$  is the observed visibility at a weather station “ $i$ ” during an hour “ $j$ ”,  $V_{\text{predicted}(i,j)}$  is estimated visibility at the same weather station “ $i$ ” during the same hour “ $j$ ”, and,  $N$  is the total number of hours.

## 3. Results

The developed models are discussed next.

### 3.1. Visibility Less than 15,000 m

OLS and WLS regression models were developed by considering all the samples with visibility value less than 15,000 m. The first model was developed using all the data. Other four models were developed for different elevation ranges (<50 m, 50 m to 250 m, 250 m to 750 m, and >750 m). The results from OLS models (**Table 1**) indicate that elevation, cloud cover, and the amount of precipitation are negatively associated with visibility less than 15,000 m when all the data are used for modeling. However, wind speed at 10-m above ground level (m10wspd) and dew point depression (tair\_dew) are positively associated with visibility less than 15,000 m when all the data are used for modeling. The positive coefficient for wind speed at 10-m indicates that the visibility increases by ~230 m for every 1 m/s increase in wind speed at 10-m. This could be attributed to boundary-layer mixing during higher wind speeds, resulting in reduced humidity leading to good visibility condition. Likewise, rainfall in the past three hours, three to six hours, and twelve to twenty-four hours are positively associated with the visibility when all the data are used for modeling. However, the rainfall during the past six to twelve hours was observed to be negatively associated with the visibility when all the data are used for modeling. In other words, if rainfall occurred during past six hours to twelve hours, the chances of lower visibility would be higher.

For regression models based on different elevation ranges (OLS and WLS), cloud cover and the amount of precipitation are negatively associated with the visibility. The results indicate that the coefficient of cloud cover is lower in the

**Table 1.** OLS and WLS regression model coefficients for visibility data < 15,000 m.

Variable	OLS					WLS				
	All	Elevation				All	Elevation			
		<50 m	50 - 250 m	250 - 750 m	>750 m		<50 m	50 - 250 m	250 - 750 m	>750 m
Elevation	-0.76	-6.79	1.58	-0.62	-2.74	-0.78	-7.59	1.50	-0.67	-3.05
Cloud cover	-32.07	-30.46	-31.21	-35.15	-40.11	-32.38	-30.43	-31.66	-35.77	-41.65
m10wspd	231.90	265.33	273.84	250.82	120.49	234.85	262.50	274.46	246.38	122.74
Precipitation	-183.39	-219.76	-159.00	-185.38	-74.06	-184.56	-228.52	-166.44	-192.18	-85.40
tair_dew	268.18	199.31	298.13	363.39	277.89	265.58	191.97	296.01	362.05	285.38
T0am-4am	10,148.4	10,084.1	9843.0	10,214.8	11,820.3	10,247.4	10,211.0	9986.0	10,350.7	12,215.9
T4am-8am	9745.7	9540.6	9519.0	9914.4	11,302.4	9833.0	9641.0	9643.3	10,042.8	11,632.0
T8am-12pm	9292.9	9106.8	9008.2	9474.8	10,865.8	9367.3	9184.1	9091.5	9588.6	11,157.9
T12pm-4pm	9639.2	9356.1	9461.5	9735.2	11,275.4	9700.7	9426.4	9544.3	9856.9	11,624.3
T4pm-8pm	9190.0	9175.7	8819.9	9300.6	10,964.4	9199.8	9239.3	8901.7	9404.0	11,274.4
T8pm-12am	9520.6	9582.6	9094.7	9528.1	11,274.1	9578.3	9672.9	9186.1	9629.3	11,585.4
R0-3hrs	711.86	679.21	557.05	753.21	1212.77	683.66	652.33	545.01	762.64	1288.2
R3-6hrs	93.73	102.47	89.48	69.91	-	74.74	99.86	84.92	50.92	-
R6-12hrs	-44.87	159.78	-118.56	-273.55	-185.96	-48.45	159.26	-157.62	-324.04	-201.17
R12-24hrs	50.15	166.28	-	-	174.51	45.96	166.97	-	-	203.30
water	-	114.83	-46.65	133.37	-	-	97.51	-67.99	115.88	-
No. of Observations	685,375	244,448	259,202	158,126	23,599	685,375	244,448	259,202	158,126	23,599
R-Squared	0.86	0.87	0.86	0.84	0.80	0.99	0.99	0.98	0.98	0.95
Adj. R-Squared	0.86	0.87	0.86	0.84	0.80	0.99	0.99	0.98	0.98	0.95
AIC	13,200,000	4,674,395	4,983,872	3,038,998	451,866	11,200,000	366,836	4,468,495	2,684,371	415,281
RMSE	3567.00	3436.30	3622.50	3606.30	3479.40	825.40	808.27	1340.60	1175.10	1602.80

Note: All the variables are significant at a 95% confidence level.

regression models based on an elevation greater than 750 m compared to the regression models based on an elevation less than 50 m. In other words, for a particular percent of cloud cover, the visibility is lower in higher elevation area (>750 m) compared to the lower elevation area (<50 m), if all the other predictor variables are kept unchanged.

In addition, the coefficient of precipitation is higher in the regression models based on an elevation greater than 750 m compared to the regression models based on an elevation less than 50 m. Also, the coefficient of precipitation is steadily increasing in the regression models based on an elevation between 50 m to 250 m and between 250 m to 750 m. Therefore, for a particular amount of precipitation, the visibility is higher in the elevation greater than 750 m compared to the elevation lower than 50 m, if all the other predictor variables are

kept constant. Further, rainfall in the past six to twelve hours is negatively associated with the visibility in regression models for elevation between 50 m to 250 m, 250 m to 750 m and greater than 750 m.

The coefficients of predictor variables in OLS and WLS models are fairly consistent. In terms of goodness-of-fit, both OLS and WLS models are acceptable, but the developed WLS models have slightly higher R-square and adjusted R-square values and lower AIC and RMSE values compared to OLS models. Similar observations were observed in case of developed models for different elevations.

### 3.2. Visibility Less than 10,000 m

The developed OLS and WLS regression models for visibility less than 10,000 m are summarized in **Table 2**. The WLS regression models outperformed the OLS

**Table 2.** OLS and WLS regression model coefficients for visibility data < 10,000 m.

Variable	OLS					WLS				
	All	Elevation				All	Elevation			
		<50 m	50 - 250 m	250 - 750 m	>750 m		<50 m	50 - 250 m	250 - 750 m	>750 m
Elevation	-0.39	-	1.89	-0.05	-0.83	-0.44	-	2.01	-0.06	-0.79
Cloud cover	-27.90	-28.20	-26.92	-28.91	-33.01	-28.39	-28.87	-27.63	-29.73	-34.03
m10wspd	171.60	181.64	205.36	192.98	132.37	178.65	189.37	211.70	199.64	138.62
Precipitation	-72.42	-99.64	-63.34	-57.13	-	-74.99	-102.16	-64.49	-58.03	-
tair_dew	126.20	107.67	147.31	137.40	117.39	127.14	102.90	145.39	134.62	116.94
T0am-4am	7060.1	6840.5	6707.1	7170.2	7597.8	7109.3	6911.7	6759.3	7240.6	7627.6
T4am-8am	6839.6	6560.4	6551.5	6942.9	7376.7	6897.9	6596.7	6613.1	7019.9	7960.1
T8am-12pm	6693.7	6488.5	6363.1	6759.9	7093.9	6722.7	6519.7	6390.5	6816.7	7080.8
T12pm-4pm	7027.4	6890.6	6717.8	6976.3	7335.9	7071.5	6940.1	6769.0	7043.1	7341.0
T4pm-8pm	6878.6	6790.8	6444.5	6945.6	7193.2	6901.1	6827.1	6468.1	7003.2	7193.1
T8pm-12am	6979.2	6930.5	6504.8	7039.2	7388.9	7012.0	6965.0	6549.9	7098.8	7405.5
R0-3hrs	916.54	1002.32	801.65	845.83	1130.51	926.88	1021.40	812.61	849.06	1176.7
R3-6hrs	158.21	184.84	155.20	126.25	-	143.17	190.18	141.72	123.49	-
R6-12hrs	50.85	199.27	46.63	-148.63	-102.69	55.87	208.90	41.51	-178.46	-95.62
R12-24hrs	-	169.87	-41.18	-119.56	-	-	177.17	-50.67	-110.93	-
Water	-222.53	-219.11	-224.95	-	-	-234.49	-219.00	-241.94	-	-
No. of Observations	361,182	117,233	137,447	90,637	15,865	361,182	117,233	137,447	90,637	15,865
R-Squared	0.83	0.84	0.83	0.81	0.79	0.98	0.98	0.97	0.97	0.96
Adj. R-Squared	0.83	0.84	0.83	0.81	0.79	0.98	0.98	0.97	0.97	0.96
AIC	6,701,768	2,173,896	2,550,679	1,681,680	292,863	5,254,038	1,846,543	2,292,924	1,485,337	264,234
RMSE	2587.90	2572.60	2590.90	2586.30	2466.40	348.78	636.84	1014.50	875.56	1000.50

Note: All the variables are significant at a 95% confidence level.

regression models based on the R-square, Adjusted R-square, AIC, and RMSE values. The elevation, cloud cover, amount of precipitation, and the presence of water are negatively associated with visibility when all the data are used for modeling. However, the elevation is observed to be positively associated with visibility for regression models based on an elevation between 50 m to 250 m. Further, in all the regression models, wind speed at 10-m and dewpoint depression are positively associated with the visibility. Also, rainfall in the past three hours, three to six hours, and twelve to twenty-four hours are positively associated with the visibility when all the data are used for modeling. However, rainfall in the past six to twelve hours and twelve to twenty-four hours are negatively associated with the visibility for regression models based on an elevation between 250 m to 750 m. For regression models based on different elevation ranges (OLS and WLS), the coefficient of cloud cover decreases with an increase in the elevation, while the coefficient of precipitation increases with an increase in the elevation.

### **3.3. Visibility Less than 5000 m**

The developed OLS and WLS regression models for the visibility less than 5000 m are summarized in **Table 3**. The cloud cover, amount of precipitation, dew point depression, and the presence of water are negatively associated with visibility when all the data are used for modeling. However, except rainfall in the past twelve to twenty-four hours, all other rainfall categories are positively associated with the visibility when all the data are used for modeling. In all the regression models, wind speed at 10-m is positively associated with visibility. However, rainfall in the past six to twelve hours and twelve to twenty-four hours are negatively associated with the visibility for regression models based on an elevation between 250 m to 750 m. For regression with different elevation ranges, the coefficient of cloud cover is lower in the regression models based on an elevation greater than 750 m compared to the regression models based on elevation less than 50 m. Likewise, dewpoint depression is negatively associated with the visibility in all the regression models except for the regression model based on elevation greater than 750 m. Also, the presence of water is negatively associated with visibility in all the regression models except for the regression model based on an elevation greater than 750 m.

### **3.4. Visibility Less than 2000 m**

The developed OLS and WLS regression models for visibility less than 2000 m are summarized in **Table 4**. The elevation, cloud cover, and the presence of water bodies are negatively associated with visibility when all the data are used for modeling. However, elevation is positively associated with the visibility for regression models based on an elevation between 50 m to 250 m and greater than 750 m. In all the regression models, wind speed at 10-m and the amount of precipitation are positively associated with the visibility. Also, the coefficient of



**Table 3.** OLS and WLS regression model coefficients for visibility data < 5000 m.

Variable	OLS					WLS				
	All	Elevation				All	Elevation			
		<50 m	50 - 250 m	250 - 750 m	>750 m		<50 m	50 - 250 m	250 - 750 m	>750 m
Elevation	0.06	5.90	1.54	–	–	0.09	6.32	1.64	–	–
Cloud cover	–17.87	–17.12	–17.33	–19.63	–24.52	–18.76	–18.45	–18.51	–20.39	–25.13
m10wspd	109.61	101.52	137.48	130.29	114.44	112.19	105.74	143.10	137.30	118.44
Precipitation	–6.62	–19.51	–	–	–	–7.87	–22.74	–	–	–
tair_dew	–14.59	–12.94	–8.24	–20.46	54.87	–14.87	–15.98	–11.51	–19.27	52.86
T0am-4am	3956.1	3542.0	3744.7	4333.0	4355.2	4006.4	3626.5	3809.6	4381.2	4400.9
T4am-8am	3795.4	3406.1	3586.0	4161.0	4224.3	3866.9	3465.5	3644.3	4202.7	4259.3
T8am-12pm	3730.9	3438.1	3499.5	3990.9	4140.0	3781.9	3504.7	3554.5	4041.3	4127.1
T12pm-4pm	4038.8	3812.2	3797.8	4257.2	4293.6	4085.9	3894.7	3883.8	4308.1	4332.6
T4pm-8pm	4225.0	3916.4	3986.9	4515.6	4382.6	4295.4	4011.9	4067.1	4578.4	4393.0
T8pm-12am	4179.3	3928.5	3894.2	4462.2	4457.4	4247.5	4022.4	3977.9	4523.7	4496.3
R0-3hrs	696.22	754.85	612.67	658.82	879.57	735.63	814.40	647.13	697.50	928.77
R3-6hrs	198.93	193.97	211.37	202.77	–	206.85	200.41	217.62	204.66	–
R6-12hrs	62.29	196.46	53.15	–69.18	–	55.34	202.04	56.21	–71.68	–
R12-24hrs	–45.67	94.78	–61.42	–179.96	–	–47.05	103.02	–68.91	–197.12	–
Water	–231.30	–308.40	–127.30	–113.81	–	–249.03	–344.62	–135.75	–134.60	–
No. of Observations	163,213	48,951	60,442	44,703	9117	163,213	48,951	60,442	44,703	9117
R-Squared	0.79	0.79	0.81	0.80	0.77	0.97	0.95	0.96	0.96	0.96
Adj. R-Squared	0.79	0.79	0.81	0.80	0.77	0.97	0.95	0.96	0.96	0.96
AIC	2,845,300	852,055	1,053,069	778,812	159,183	2,515,898	772,895	936,818	696,850	140,030
RMSE	1476.70	1456.90	1469.00	1468.30	1496.00	538.30	649.02	562	587.06	523.29

Note: All the variables are significant at a 95% confidence level.

precipitation is higher in the regression models based on an elevation greater than 750 m compared to the regression models based on an elevation less than 50 m. On the other hand, dewpoint depression is negatively associated with the visibility for regression models based on an elevation between 50 m to 250 m and positively associated with the visibility for regression models based on an elevation greater than 750 m. In addition, the rainfall in the past twelve to twenty-four hours is negatively associated with the visibility for regression models based on an elevation between 250 m to 750 m and greater than 750 m. Also, the presence of water is negatively associated with the visibility in the majority of the regression models. In addition, the coefficient of rainfall in the past six to twelve hours is lower compared to the earlier hours after the rainfall. In other words,

**Table 4.** OLS and WLS regression model coefficients for visibility data < 2000 m.

Variable	OLS					WLS				
	All	Elevation				All	Elevation			
		<50 m	50 - 250 m	250 - 750 m	>750 m		<50 m	50 - 250 m	250 - 750 m	>750 m
Elevation	-0.13	-0.58	0.49	-0.19	0.85	-0.14	-0.63	0.44	-0.19	0.93
Cloud cover	-2.78	-2.66	-2.88	-3.39	-34.66	-2.92	-2.72	-3.09	-3.70	-36.16
m10wspd	33.65	30.14	40.18	30.80	56.38	35.10	29.98	41.39	32.24	61.23
Precipitation	6.44	3.38	5.78	13.54	36.39	6.89	4.29	6.98	14.34	40.65
tair_dew	-	-4.90	-	-	66.70	-	-3.75	-	-	65.18
T0am-4am	1133.3	1106.0	1057.3	1268.9	3271.0	1145.8	1104.0	1078.8	1296.3	3307.3
T4am-8am	1090.6	1065.4	1023.2	1208.8	3317.7	1103.3	1069.2	1041.7	1239.8	3362.6
T8am-12pm	1001.4	992.1	924.2	1112.8	3220.3	1010.3	995.7	947.9	1134.0	3274.9
T12pm-4pm	988.3	961.2	898.9	1117.5	3196.3	992.6	961.3	912.9	1143.8	3239.6
T4pm-8pm	1153.3	1132.6	1069.0	1308.9	3252.7	1174.6	1137.2	1099.2	1359.1	3304.6
T8pm-12am	1146.8	1153.9	1025.4	1312.3	3234.7	1157.5	1160.0	1050.0	1346.7	3276.0
R0-3hrs	137.14	102.13	110.72	156.78	219.15	135.92	102.04	111.51	159.57	250.79
R3-6hrs	58.60	75.76	45.23	75.17	-	57.59	77.53	44.61	74.71	-
R6-12hrs	18.94	30.46	21.71	-	-	17.93	29.65	19.22	-	-
R12-24hrs	-	42.75	-	-52.69	-57.10	-	42.13	-	-55.72	-49.05
Water	-41.79	-47.31	-	-66.86	-	-45.20	-47.48	-	-76.24	-
No. of Observations	54,232	17,125	19,144	14,590	3373	54,232	17,125	19,144	14,590	3373
R-Squared	0.76	0.78	0.77	0.76	0.66	0.98	0.97	0.96	0.96	0.93
Adj. R-Squared	0.76	0.78	0.77	0.76	0.66	0.98	0.97	0.96	0.96	0.93
AIC	821,396	258,336	289,282	221,156	51,702	665,843	222,987	249,388	192,932	45,662
RMSE	470.56	456.35	462.27	473.24	514.49	112.14	162.58	163	179.89	210.16

Note: All the variables are significant at a 95% confidence level.

the chances of low visibility condition are higher between six to twelve hours after the rainfall compared to the first six hours after the rainfall.

Based on the mode or type of transportation application, the developed regression models would be helpful to understand the relationship between predictor variables and visibility. The WLS model for visibility less than 15,000 m would be suitable for air traffic or similar operations, while the WLS for visibility less than 2000 m would be suitable to share visibility information with motorists. In addition, both the aforementioned models are observed to be best-fitted models with higher R-square and adjusted R-square values and lower AIC, and lower RMSE values.

### 3.5. Validation of the Selected Model

The WLS regression model for visibility less than 15,000 m and the WLS regres-

sion model for visibility less than 2000 m by considering the complete dataset were validated with the separate data acquired from the weather stations. The results indicate that the majority of the samples fall under the visibility range of 15,000 m to 10,000 m with MAPE values between 10% and 30% (Figure 1). In addition, the majority of the samples fall under the MAPE greater than 40% category for the regression model based on visibility less than 2000 m (Figure 2).

#### 4. Conclusions

This study focuses on identifying predictor variables associated with different visibility and elevation ranges. Based on the application/mode of transportation,

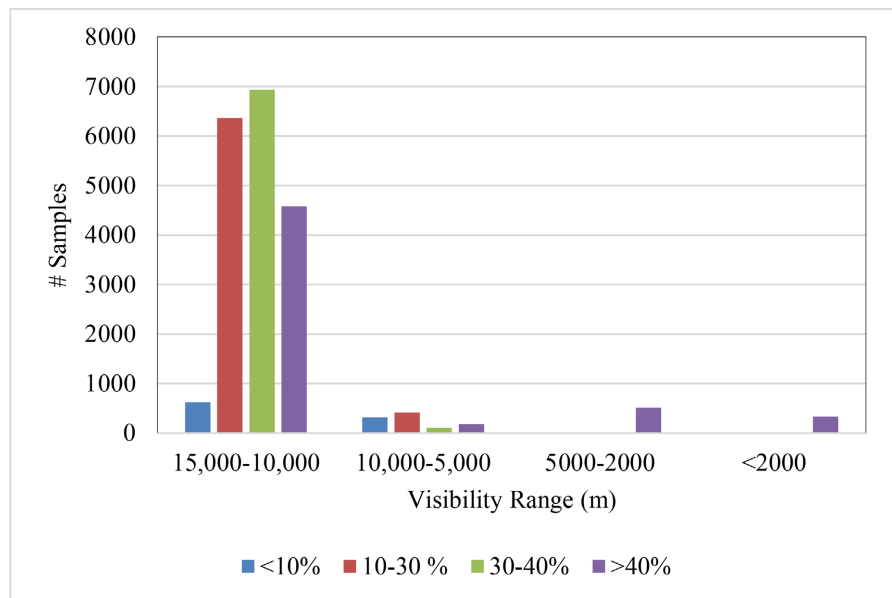


Figure 1. Distribution of errors—visibility less than 15,000 m.

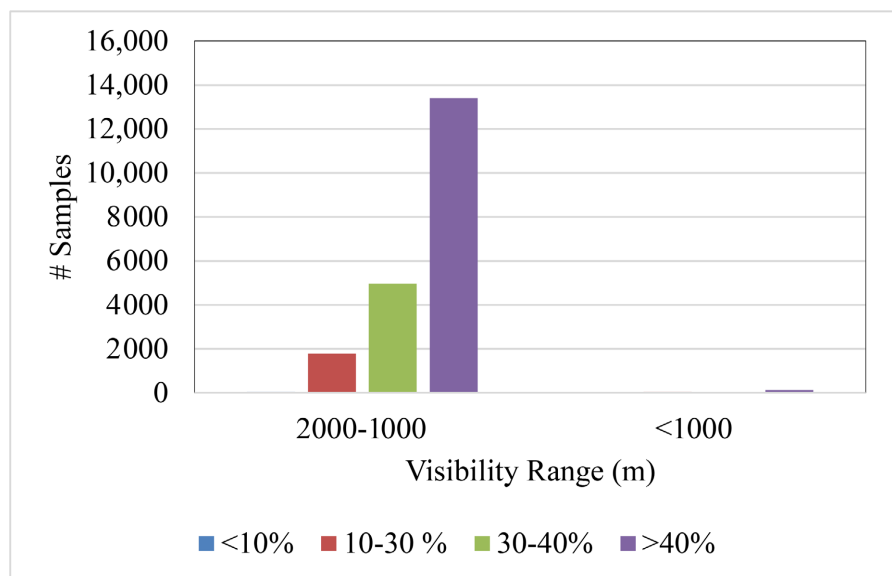


Figure 2. Distribution of errors—visibility less than 2000 m.

the developed models could be used to identify the effect of meteorological variables on visibility. The WLS regression model for visibility less than 2000 m by considering all the samples and irrespective of elevation is the best-fitted model for road traffic operations and safety. Typically, the presence of water within the vicinity contributes to low visibility conditions (less than 2000 m). Also, the contribution of cloud cover on the low visibility conditions increases with an increase in the elevation if all other predictor variables are kept constant. In general, the chances of low visibility condition are higher between six to twelve hours after the rainfall when compared to the first six hours after the rainfall.

The WLS regression model for visibility less than 15,000 m, by considering all the samples and irrespective of elevation, is the best-fitted model for air traffic operations and safety. For visibility less than 15,000 m, the contribution of cloud cover on visibility increases with an increase in the elevation while the influence of precipitation on visibility decreases with an increase in the elevation. Also, the chances of visibility less than 15,000 m are higher between six to twelve hours after the rainfall when compared to the first six hours after the rainfall.

Based on the findings, implementing dynamic message sign-boards/communicating the information through radio/phones or the Internet to the motorists in the mountainous areas, near the water bodies and between six to twelve hours after the rainfall about the possibility of low visibility condition could improve the safety for motorists.

Comparing the visibility from weather stations, numerical models, satellite data, and for regions with different climatic and topographical conditions warrant further investigation.

### **Acknowledgments**

The contents of this paper reflect the views of the authors and not necessarily the views of the University of North Carolina at Charlotte (UNC Charlotte), the University of North Carolina at Asheville (UNC Asheville) or the NCDOT. The authors are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of either UNC Charlotte, UNC Asheville, NCDOT or the Federal Highway Administration (FHWA) at the time of publication. This paper does not constitute a standard, specification, or regulation.

Special thanks are extended to Christopher J. Oliver, Jason Holmes, Jimmy Hamrick, George D. Eckart, Meredith M. McDiarmid, and Ernest Morrison of NCDOT for providing excellent support, guidance, and valuable inputs for successful completion of this project.

### **Disclaimer**

This paper is disseminated in the interest of information exchange. The views, opinions, findings, and conclusions reflected in this paper are the responsibility of the authors only and do not represent the official policy or position of the

University of North Carolina at Charlotte or other entity. The authors are responsible for the facts and the accuracy of the data presented herein. This paper does not constitute a standard, specification, or regulation.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Fultz, A.J. and Ashley, W.S. (2016) Fatal Weather-Related General Aviation Accidents in the United States. *Physical Geography*, **37**, 291-312. <https://doi.org/10.1080/02723646.2016.1211854>
- [2] United States Department of Transportation. Federal Highway Administration (FHWA) (2018) How Do Weather Events Impact Roads? [https://ops.fhwa.dot.gov/weather/q1\\_roadimpact.htm](https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm)
- [3] Oliver, C.J. (2013) Statewide Study of Fog Related Crashes in North Carolina. Traffic Safety Unit Transportation Mobility and Safety Division, North Carolina Department of Transportation. <https://connect.ncdot.gov/resources/safety/Crash%20Data%20and%20TEAAS%20System/Crash%20Data%20and%20Information/Statewide%20Study%20of%20Fog%20Related%20Crashes%20in%20NC.pdf>
- [4] Mathew, S. and Pulugurtha, S.S. (2022) Quantifying the Effect of Rainfall and Visibility Conditions on Road Traffic Travel Time Reliability. *Weather, Climate, and Society*, **14**, 507-519. <https://doi.org/10.1175/WCAS-D-21-0053.1>.
- [5] Vislocky, R.L. and Fritsch, J.M. (1997) An Automated, Observations-Based System for Short-Term Prediction of Ceiling and Visibility. *Weather and Forecasting*, **12**, 31-43. [https://doi.org/10.1175/1520-0434\(1997\)012%3C0031:AAOBSF%3E2.0.CO;2](https://doi.org/10.1175/1520-0434(1997)012%3C0031:AAOBSF%3E2.0.CO;2)
- [6] Hilliker, J.L. and Fritsch, J.M. (1999) An Observations-Based Statistical System for Warm-Season Hourly Probabilistic Forecasts of Low Ceiling at the San Francisco International Airport. *Journal of Applied Meteorology*, **38**, 1692-1705. [https://doi.org/10.1175/1520-0450\(1999\)038%3C1692:AOBSSF%3E2.0.CO;2](https://doi.org/10.1175/1520-0450(1999)038%3C1692:AOBSSF%3E2.0.CO;2)
- [7] Gultepe, I., Tardif, R., Michaelides, S.C., Cermak, J., Bott, A., Bendix, J., Müller, M.D., Pagowski, M., Hansen, B., Ellrod, G., Jacobs, W. and Cober, S.G. (2007) Fog Research: A Review of Past Achievements and Future Perspectives. *Pure and Applied Geophysics*, **164**, 1121-1159. <https://doi.org/10.1007/s00024-007-0211-x>
- [8] Meyer, M.B., Jiusto, J.E. and Lala, G.G. (1980) Measurements of Visual Range and Radiation-Fog (Haze) Microphysics. *Journal of Atmospheric Sciences*, **37**, 622-629. [https://doi.org/10.1175/1520-0469\(1980\)037%3C0622:MOVRAR%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1980)037%3C0622:MOVRAR%3E2.0.CO;2)
- [9] Jiusto, J.E. (1981) Fog Structure. In: Hobbs, P.V. and Deepak, A., Eds., *Clouds: Their Formation, Optical Properties and Effects*, Academic Press, Cambridge, 187-239. <https://doi.org/10.1016/B978-0-12-350720-4.50009-0>
- [10] Tardif, R. and Rasmussen, R.M. (2008) Process-Oriented Analysis of Environmental Conditions Associated with Precipitation Fog Events in The New York City Region. *Journal of Applied Meteorology and Climatology*, **47**, 1681-1703. <https://doi.org/10.1175/2007JAMC1734.1>
- [11] Pulugurtha, S.S., Mane, A.S., Duddu, V.R. and Godfrey, C.M. (2019) Investigating the Influence of Contributing Factors and Predicting Visibility at Road Link-Level. *Heliyon*, **5**, e02105. <https://doi.org/10.1016/j.heliyon.2019.e02105>

- [12] Del Greco, S.A., Lott, N., Hawkins, K., Baldwin, R., Anders, D.D., Ray, R., Dellinger, D., Jones, P. and Smith, F. (2006) Surface Data Integration at NOAA's National Climatic Data Center: Data Format, Processing, QC, and Product Generation. *22nd International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*, Atlanta, 29 January-2 February 2006, 4 p. <http://ams.confex.com/ams/pdfpapers/100500.pdf>
- [13] Smith, A., Lott, N. and Vose, R. (2011) The Integrated Surface Database: Recent Developments And Partnerships. *Bulletin of the American Meteorological Society*, **92**, 704-708. <https://www.jstor.org/stable/26218543>  
<https://doi.org/10.1175/2011BAMS3015.1>
- [14] Lott, J.N. (2004) The Quality Control of the Integrated Surface Hourly Database. *14th Conference on Applied Climatology*, Seattle, 11-15 January 2004, 7 p. <https://ams.confex.com/ams/pdfpapers/71929.pdf>
- [15] Godfrey, C.M. (2015) Improved Climatic Data for Mechanistic-Empirical Pavement Design. Report FHWA/NC/2014-01, North Carolina Department of Transportation. <https://connect.ncdot.gov/projects/research/RNAProjDocs/2014-01FinalReport.pdf>
- [16] Glickman, T.S. (2000) Glossary of Meteorology. 2nd Edition, American Meteorological Society, Boston, 850 p.
- [17] Gujarati, D.N. and Porter, D.C. (2009) Basic Econometrics. 5th Edition, McGraw-Hill, New York.