Safety Science 119 (2019) 21–28

Contents lists available at ScienceDirect

Safety Science

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

Complementary methodologies to identify weather conditions in naturalistic driving study trips: Lessons learned from the SHRP2 naturalistic driving study & roadway information database

Ali Ghasemzadeh[a,*], Britton E. Hammit[b,a], Mohamed M. Ahmed[a], Hesham Eldeeb[c]

[a] University of Wyoming, Department of Civil & Architectural Engineering, 1000 E University Ave, Dept. 3295, Laramie, WY 82071, United States
[b] FHWA Turner Fairbank Highway Research Center, 6300 Georgetown Pike, McLean, VA 22101, United States
[c] Deep Numerics, LLC, 3317 Curving Oaks Way, Orlando, FL 32820, United States

ARTICLE INFO

Keywords:
Naturalistic driving data reduction
SHRP2
Adverse weather conditions
Data visualization
Visibility identification

ABSTRACT

Adverse weather conditions play a considerable role in the safety and efficiency of the transportation network. Many studies have aimed to quantify the impact that different weather conditions have on transportation safety and mobility; however, most studies have evaluated the network capacity, average speed, and other macroscopic measures without capturing specific driving characteristics. In order to understand specific driving behavior and performance characteristics that exist during different environmental conditions, high resolution vehicle data and video footage are required. The SHRP2 sponsored the generation of a large Naturalistic Driving Study (NDS) database – which provides vehicle time series data, front and rear video, driver video, external sensor readings, and driver surveys – and the Roadway Information Database (RID) – which is a complementary database with geospatial data for commonly driven roads in the NDS and other ancillary data sources, including annual traffic, roadway geometry, accident reports, weather conditions, and 511 alerts. The purpose of this study is to leverage these SHRP2 databases and weather data from the National Climatic Data Center (NCDC) to extract trips that occur during adverse weather conditions. The extraction of weather-related trips from a NDS is unprecedented, and this study presents three complementary methodologies used in parallel to acquire relevant trips from the SHRP2 NDS database. A semi-automated data reduction procedure was developed to process the raw trip files into a format that further analysis and modeling could be completed. This novel approach to NDS trip acquisition and reduction could be extended to other naturalistic driving studies worldwide.

1. Introduction

Adverse weather conditions are one of the critical factors impacting the safety and mobility of roadways (Ghasemzadeh and Ahmed, 2017a, 2017b). Many previous studies have aimed to quantify the impact adverse weather conditions have on the transportation network; for example, several studies concluded that injury and property damage only (PDO) crash rates increased during snow events (Eisenberg and Warner, 2005; Knapp et al., 2000; Ghasemzadeh and Ahmed, 2018a, 2018b). While most studies have resolved a significant difference in safety and mobility during adverse weather, the results of these studies are inconclusive as the findings are not consistent in magnitude (Ghasemzadeh and Ahmed, 2018c; Ghasemzadeh et al., 2018; Hammit et al., 2018). Several studies concluded that crashes increase by 100% or more due to vision obstruction during rainfall (Brodsky and Hakkert, 1988), while others found more moderate, but still statistically significant, increases (Andreescu and Frost, 1998; Andrey and Olley, 1990). Studies have also focused on the impact reduced visibility has on driving speed selection during different weather conditions, such as rain and snowfall (Ahmed and Ghasemzadeh, 2018, 2017; Kyte et al., 2001; Maze et al., 2006; Ahmed et al., 2017).

Based on the National Highway Traffic Safety Administration (NHTSA) data, approximately 22% of crashes are weather-related each year, leading to over 6000 fatalities, and more than 445,000 injuries (Hamilton, 2016). Current records indicate that most weather-related crashes occur during rainfall and wet surface conditions (46% during rainfall and 73% on wet pavement), followed by winter conditions: snow or sleet (17%), icy pavement (13%), snowy or slushy pavement (14%) (Hamilton, 2016). Booz Allen Hamilton showed that 19% of injury crashes and 16% of fatal crashes are weather related (Hamilton, 2016). The recently published 2016 Highway Capacity Manual (HCM) has discussed the impact of inclement weather conditions on traffic

* Corresponding author.
E-mail addresses: aghasemz@uwyo.edu (A. Ghasemzadeh), mahmed@uwyo.edu (M.M. Ahmed).

https://doi.org/10.1016/j.ssci.2019.01.006
Received 1 January 2018; Received in revised form 25 September 2018; Accepted 8 January 2019
Available online 25 January 2019
0925-7535/

operation and has provided Weather Adjustment Factors (WAFs) based on (1) weather type and intensity and (2) facility free-flow speed (FFS). As an example, considering a freeway corridor with a 65 mph FFS, freeway capacity reductions are predicted to be 8% and 14% for medium rain and heavy rain, respectively (Highway Capacity Manual, 2016).

1.1. Characterizing driving behavior and performance in adverse weather conditions

Many studies have focused on quantifying the operational and safety impacts that adverse weather conditions have on various roadway types by evaluating macroscopic traffic flow data. Most of these studies have utilized aggregated traffic and weather data (e.g., average speed, headway, and global weather information) to formulate conclusions defining the impact of weather on network operation and safety. Such studies utilized traffic and weather data collected from inductive loop detectors (iLD), Automatic Vehicle Identification (AVI) systems, and Roadway Weather Information System (RWIS) to distinguish 'crash prone' conditions from 'normal' conditions (Ahmed et al., 2014; El-tawab and Olariu, 2010). Few studies, however, have attempted to identify fundamental characteristics of driving behavior and performance that cause noticeable speed and flow reductions (Kamrani et al., 2018a, 2018b, 2018c). Individual driver characteristics, preferences, and tendencies have a heavy influence on traffic flow metrics, such as roadway capacity; therefore, it is crucial to understand and characterize these basic driver impacts as we discuss the network-wide impacts. Quantifying changes in driver behavior and performance is critical for many safety and operations planning procedures, such as microsimulation modeling (Das et al., 2018; Khan et al., 2017).

The approach presented in this paper offers a mechanism for acquiring and reducing disaggregate trip data available through the second Strategic Highway Research Program's (SHRP2) Naturalistic Driving Study (NDS) database and Roadway Information Database (RID).

1.2. Naturalistic driving studies

Naturalistic driving studies (NDS) are a one-of-a-kind mechanism for collecting natural driving behavior and characteristics. NDS are not set in a controlled environment, such as driver simulators, instrumented research vehicle tests, or self-reporting questionnaires; therefore, the collected data is more representative of natural driving and contains a wide variety of environment and traffic conditions. Naturalistic driving studies hire participants with personal vehicles, equip their vehicles with discreet equipment to monitor driving conditions, and simply instruct the participants to go about their normal schedules for a period of time (generally multiple months) while data is collected. This method of data collection enables researchers to better understand the driving behavior and performance illustrated by different individuals in a truly "natural" environment.

1.3. SHRP2 data description

The SHRP2 was developed to support the investigation of the fundamental causes of highway crashes and congestion by supporting focused research projects. The research projects supported by the SHRP2 are intended to identify and implement practical countermeasures with significant safety benefits through a comprehensive understanding of drivers' performance and behavior. Two databases were generated to support this goal: (1) the Naturalistic Driving Study (NDS) database and (2) the Roadway Information Database (RID).

The SHRP2 NDS database contains trips from more than 3500 drivers located in six different states (Florida, Indiana, New York, North Carolina, Pennsylvania, and Washington) (Hankey et al., 2016; Perez et al., 2016). Participants were recruited for time periods between a few months to a couple of years, and during that time, their vehicle-use was recorded continuously. Throughout this study, 5 million trip files were recorded, encompassing nearly 50 million vehicle miles, making this project one of the world's largest naturalistic driving studies. The NDS data were collected and are maintained by Virginia Tech Transportation Institute (VTTI) (Hankey et al., 2016; Perez et al., 2016). The SHRP2 RID is a geodatabase containing roadway and ancillary data complementary to the NDS data collection sites. The RID inventory data includes information related to roadway geometry, roadway features, crash histories, traffic counts, work zones, ongoing safety campaigns, and basic weather information. The quantity and quality of available ancillary data is variant for each NDS state (Smadi et al., 2014).

While the SHRP2 databases were generated with safety-based priorities, operational strategies can also be improved through the careful consideration of the dataset. Strategies for reducing fuel consumption, improving travel times, and more effectively predicting driving behavior for modeling and planning purposes can and should be produced to fully leverage the extensive databases.

1.4. Paper contribution

The SHRP2 NDS database contains a surplus of data that can be used to answer many research questions; however, the extraction of trips occurring in adverse weather conditions is complex because weather information for each trip is unknown. In this study, three complementary methodologies were developed to identify weather-related trips using the SHRP2 databases, NDS and RID, alongside ancillary weather data obtained from the National Climatic Data Center (NCDC). These innovative methodologies add value to the existing SHRP2 database by introducing weather and road condition information. By integrating these databases, the authors are now able to address research questions aiming to quantify measurable driver behavior and performance characteristics in adverse weather conditions, which are the root cause of network-wide impacts during those adverse conditions; therefore, a clear understanding at a microscopic, driver level, is the first step in identifying the best countermeasures to improve the safety and operation of the transportation network during adverse conditions.

In addition, this paper introduces a semi-automated data reduction procedure used to efficiently process a large number of SHRP2 NDS trips into a functional format. Two tools were generated for the data reduction procedure: (1) Visualization and Reduction Tool and (2) Python-based Analytic tool. The Visualization and Reduction Tool is used to synchronize the time series vehicle data, video footage, and radar data provided for each NDS trip, as well as derive the visibility index from the forward-facing video. The python-based Analytic tool is used to reduce the dimensionality of the time series data, aggregate the data into 1-minute segments, create manual observation templates, and ultimately merge the manual observation templates, roadway geometric features (from the RID), and driver survey responses into a single dataset that can be used for further analysis and modeling.

2. Data acquisition methodologies

Augmenting the SHRP2 NDS and RID databases with specific weather information is complex; therefore, three complementary methodologies were developed to achieve this goal. The procedures used to identify weather-related trips for each of the three methods is outlined in Fig. 1. Each method is described in detail in the subsequent sections.

2.1. Method 1: Wiper status

In the first method, a vehicle's front windshield wiper status (or setting) is used to identify weather-related trips. While the wiper setting can indicate precipitation intensity, the reported wiper settings are not uniform between different vehicles. The reported wiper setting

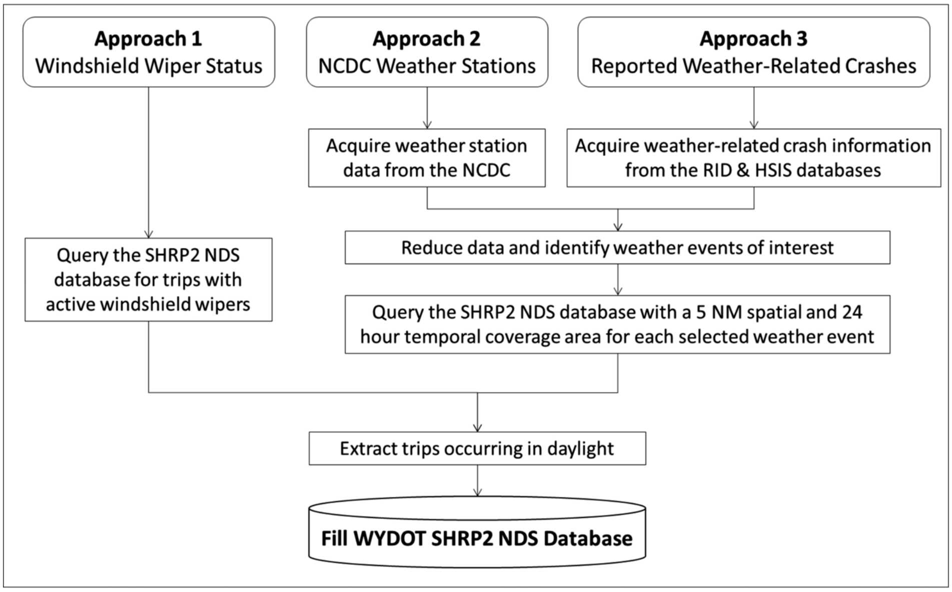


Fig. 1. Procedure for identifying weather-related NDS trips.

available through the SHRP2 NDS database indicates the position of the wiper switch rather than the actual wiper swipe rate, which is also different among different participant vehicles. Moreover, different drivers have unique tolerances and preferences impacting their selection of a wiper speed, and splashes from other vehicles affect the selection of a comfortable wiper speed.

A preliminary investigation using this data acquisition method was completed to evaluate the effectiveness of the approach. This preliminary investigation identified an issue in identifying windshield wiper use for Honda Civics: the wiper blades of Honda Civic vehicles do not cover the full windshield and the front camera was blurred due to the precipitation and deemed unusable (Ahmed et al., 2015; Ghasemzadeh and Ahmed, 2017c). From the preliminary investigation, a series of procedures were established to extract NDS trips occurring in precipitation without introducing bias to the sample data. The final NDS extraction steps for weather-related trips using the first method are listed below:

1. Only vehicles with multiple wiper settings were targeted; vehicles that did not include the full spectrum of values for the wiper status (0, 1, 2, and 3) were filtered out – as vehicles with only on/off wiper settings would not provide an indication of rain intensity.
2. Months with heavy precipitation in the SHRP2 states were targeted.
3. Trips not occurring during daylight on freeways were filtered out; freeways were considered due to the project scope and nighttime trips were eliminated because of the low video resolution.
4. Potential events were tagged with the duration of the trip that different wiper settings (0, 1, 2, and 3) were active to identify trips of interest with various levels of precipitation.

2.2. Method 2: NCDC weather stations

While the first method considered a data extraction approach using the data available directly in the NDS database, this second method branches out to leverage external data sources to capture more trips occurring during a variety of adverse weather conditions, not limited to precipitation events with active windshield wipers. For this process, weather data were extracted from the National Climate Data Center (NCDC), which is available through Climate Data Online (CDO). The NCDC archives weather data from various weather stations nationwide, including radar, satellites, airport weather stations, and military weather stations. Among these data sources, the airport weather stations proved to be the most successful in identifying adverse weather events.

Airport weather stations monitor weather conditions continuously and record the weather parameters according to predefined changes in their values; for that reason, the data do not follow a specific time pattern, but report weather conditions relative to real time weather changes. The weather parameters collected include visibility, temperature, humidity, wind speed and direction, and precipitation. Among these parameters, visibility is considered one of the most critical factors affecting driver behavior. Visibility can generally be described as the maximum distance that an object can be clearly perceived against the background sky; decreased visibility can be a result of both natural phenomena (e.g., fog, mist, haze, snow, rain, windblown dust, etc.) and human induced activities (transportation, agricultural activities, fuel combustion, etc.). Weather stations cannot directly measure visibility, but rather calculate it from a measurement of light extinction, which includes the scattering and absorption of light by particles and gases (Ahmed et al., 2012).

Previous studies concluded that airport weather stations can provide spatial-temporal weather conditions for adjacent roadways within 5 nautical miles (n.m.) and within a 2-hour time period at an accuracy between 60% and 80% (Mohamed Ahmed et al., 2014). Over 5 Gigabytes of weather data from more than 250 weather stations in the NDS states (between 2010 and 2013) were collected from the National Oceanic and Atmospheric Administration (NOAA) - National Climatic Data Center (NCDC) website (Mohamed Ahmed et al., 2014). In this study, daily weather data were acquired from the NCDC, and NDS trips were requested based on the extracted weather data to identify all trips impacted by the adverse weather events – even those impacted by ice or slush on road surfaces, not only those occurring during active precipitation or fog. Therefore, NDS trips occurring within a 5 n.m. radius relative to a weather station on days with adverse weather conditions were extracted from the SHRP2 NDS database.

Fig. 2 shows the location weather stations (airport and other weather stations) used to identify rain-related trips in Washington and the 5 n.m. coverage area (indicated by the red oval) used in the NDS trip data acquisition process.

In total, 24 GIS-shape files representing rain, snow, fog, and wind conditions for the six NDS states were provided to VTTI for the extraction of NDS trips. Due to the weather data aggregation level and the importance of protecting PII (personally identifiable information), days with large number of weather events were identified and all trips

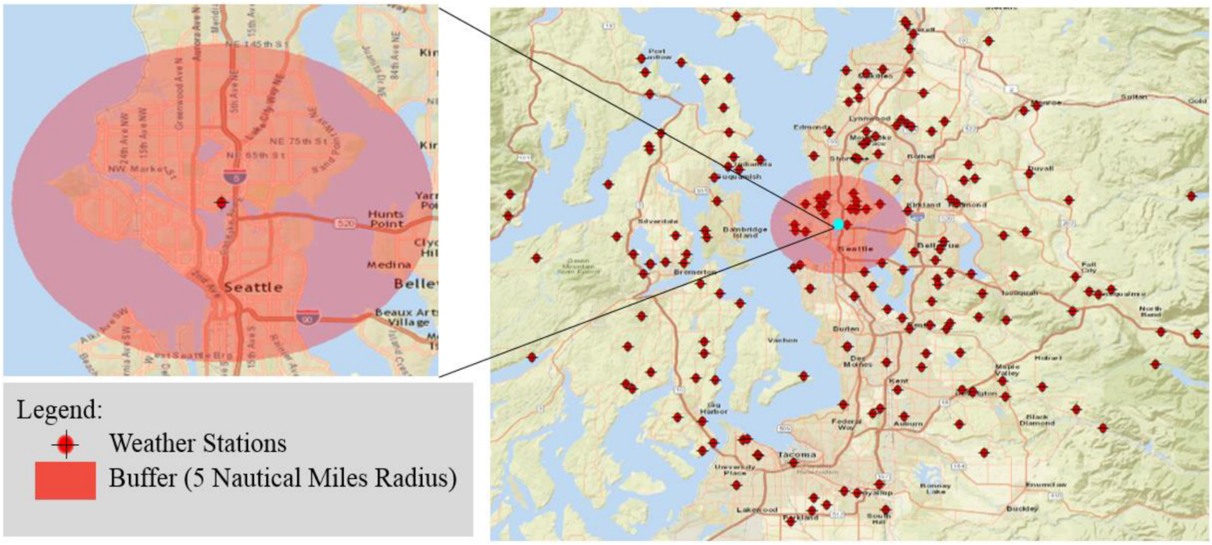


Fig. 2. Weather stations and their 5 n.m. coverage area – Washington State.

occurring within the coverage area on that day were requested. While this process produces significantly more false-positives (potential weather-related trips that are not in adverse weather conditions), it also enables the collection of trips that could not be detected using other methods.

2.3. Method 3: weather-related crashes

Similar to the second method, the third method for extracting SHRP2 NDS trips occurring in adverse weather conditions utilized the crash data provided in the RID. Using this methodology, crash locations were essentially considered as weather stations, and the weather description provided by the crash report was used to indicate the local spatial-temporal weather conditions. Crashes that occurred amidst adverse weather conditions were identified, and trips occurring on the day of the accident, on freeways, and within a 5 n.m. radius of the crash location were flagged.

3. Data reduction procedures

Once the data acquisition process was completed and the SHRP2 NDS trips were received, data reduction and processing procedures were established. Processing vehicle data, specifically those extracted from vast naturalistic driving study databases, is challenging for many reasons:

- The size and complexity of the data;
- Time consuming processing and reduction of video footage;
- Linking the continuous vehicle information and the RID;
- Identifying surrogates for increased traffic safety risks; and
- Defining baselines in normal driving conditions.

Therefore, data reduction procedures were generated to facilitate an efficient and effective procedure for processing the large trip files and their complimentary data sources. For this study, the data utilized for the analysis include: forward-facing and rear-facing videos, driver demographics, driver survey responses, and selected vehicle time series data. In addition, ancillary sources were used to complement the NDS data, including the SHRP2 RID, available images via Google Maps, and weather data extracted from the NCDC.

A semi-automated data reduction process was developed to efficiently process the large quantity of NDS trips. The procedures for this process are shown in Fig. 3. The parallelograms on the left indicate the process inputs, while the parallelograms on the right indicate the process outputs; the processes in rectangles indicate automated processes, while the obtuse trapezoid indicates a manual process.

3.1. Dimensionality reduction to 1-Min chunks

The developed semi-automated procedure reduces the dimensionality of the data by extracting relevant time series variables that were previously selected as being pertinent to characterizing driving behavior in adverse weather conditions. Preliminary studies using this data revealed a high variability in weather conditions within a single trip; therefore, each trip was divided into 1-min chunks to create homogeneous segments with similar environmental and traffic conditions (Ghasemzadeh and Ahmed, 2017c; Ghasemzadeh and Ahmed, 2018d). This process of splitting trips into 1-min chunks is a data preparation protocol introduced to enable the capability of the research team to evaluate a substantial number of trips, drivers, and conditions in the most efficient and effective manner possible.

3.2. Video observation template

The process of data reduction is semi-automatic because it is comprised of an automated analysis component and a manual analysis component. The research team created an analytic Python-based tool that automatically reduces the NDS time series data into statistics, plots, and observation templates; however, manual observation and annotation of videos to determine traffic and environmental conditions is still required. Observation templates for each trip are automatically created by the Analytic tool, identifying each 1-min chunk within the trip, its corresponding timestamps, and leaves data fields for video reviewers to annotate observed environmental and traffic conditions for the time slice. A sample of this template is provided in Fig. 4.

The template for each trip can be distinguished by the unique Event ID number provided in the second column on the top (Fig. 4). The number of samples refers to the number of 1-min chunks within the trip. For this condensed example, only 4 min of the trip were considered; therefore, 4 samples, their corresponding time stamps, and the blank locations for user input are provided. Additional instructions are added to the document, and a location for recording the video-reviewer's name, date, and additional comments are provided.

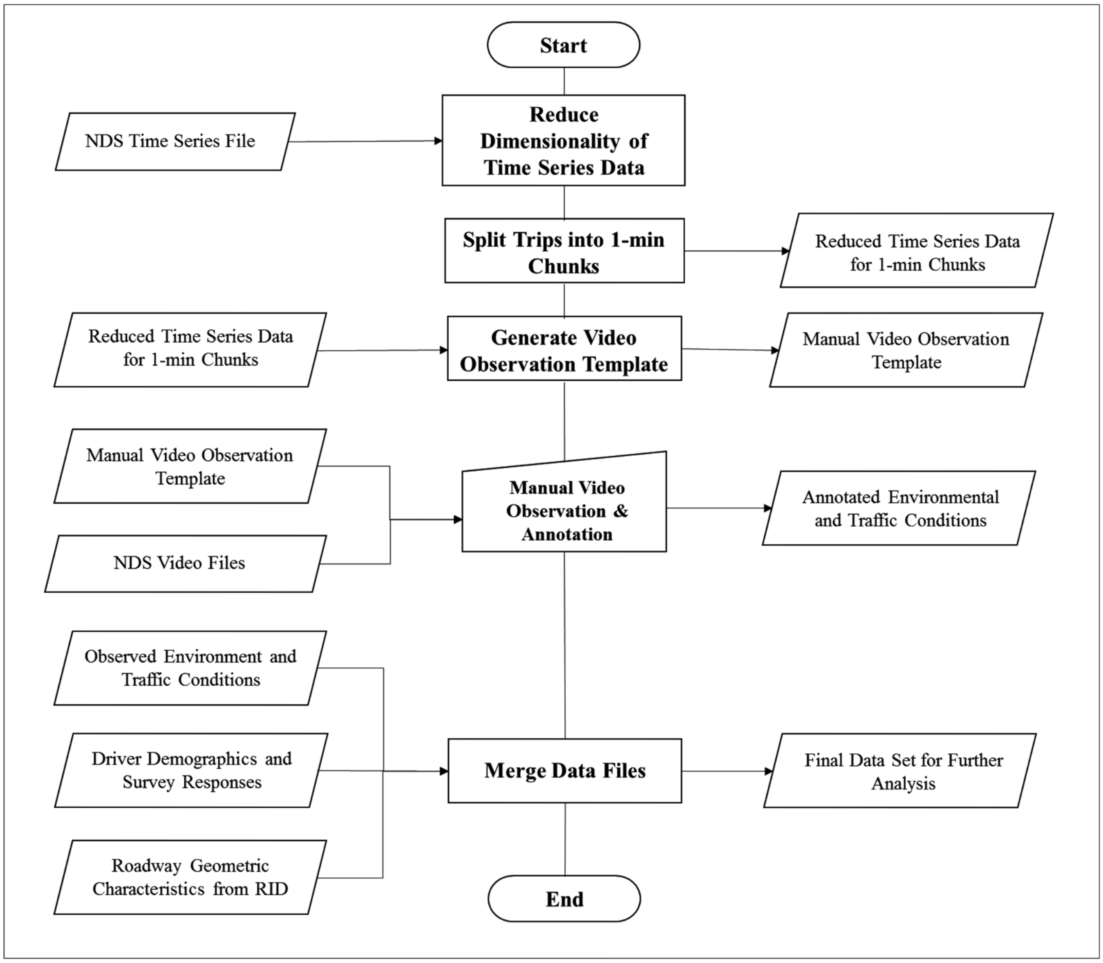
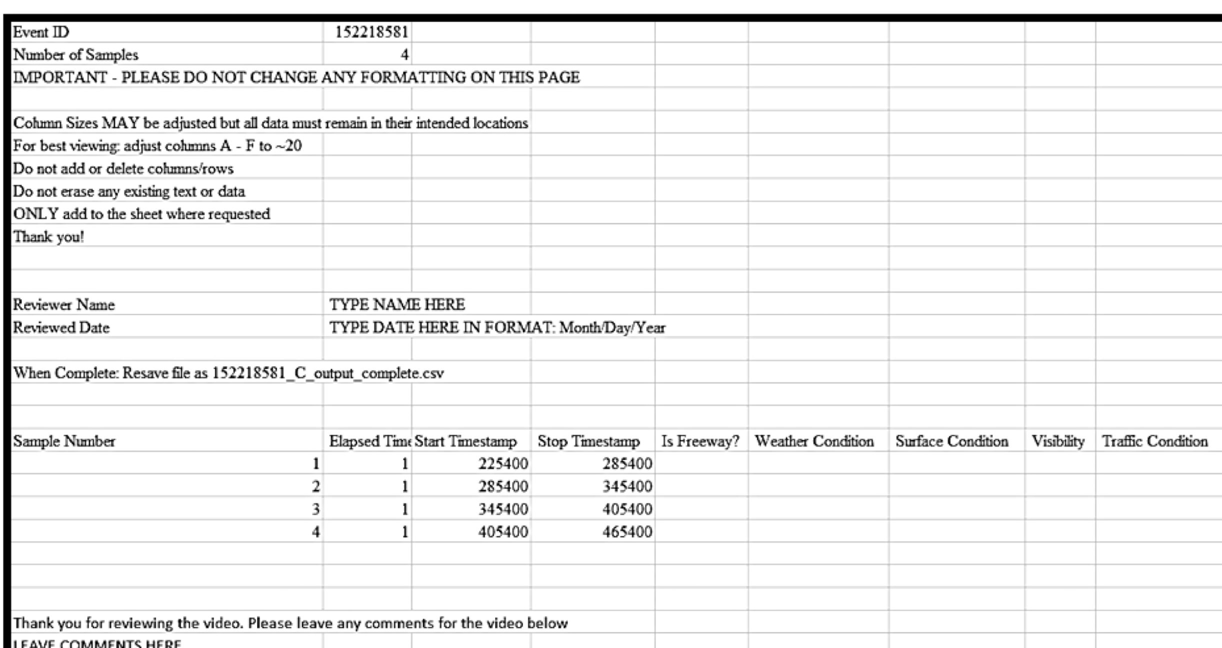


Fig. 3. Semi-automated data reduction procedures.



Event ID	152218581							
Number of Samples	4							
IMPORTANT - PLEASE DO NOT CHANGE ANY FORMATTING ON THIS PAGE								
Column Sizes MAY be adjusted but all data must remain in their intended locations								
For best viewing: adjust columns A - F to ~20								
Do not add or delete columns/rows								
Do not erase any existing text or data								
ONLY add to the sheet where requested								
Thank you!								
Reviewer Name	TYPE NAME HERE							
Reviewed Date	TYPE DATE HERE IN FORMAT: Month/Day/Year							
When Complete: Resave file as 152218581_C_output_complete.csv								
Sample Number	Elapsed Time	Start Timestamp	Stop Timestamp	Is Freeway?	Weather Condition	Surface Condition	Visibility	Traffic Condition
1	1	225400	285400					
2	1	285400	345400					
3	1	345400	405400					
4	1	405400	465400					
Thank you for reviewing the video. Please leave any comments for the video below								
LEAVE COMMENTS HERE								

Fig. 4. Sample video observation template produced by python-based analytic tool.

3.3. Manual video observation and annotation

The most time-consuming process of the data reduction is the manual video observation and annotation required to confirm the weather and traffic conditions experienced on each trip. Advanced machine learning and image processing techniques are being utilized to leverage and automate this process; however, variation in quality of the NDS data, as well as the physical location of the camera are challenging obstacles to overcome. Therefore, in the meantime, a systematic procedure is in place to gather environmental and traffic data from the video footage manually. The manual video observation and annotation is also needed to validate the automated image processing techniques.

Aside from the labor and time-intensive attributes of manual video observation, another disadvantage to completing these observations with multiple viewers is the difference in perception and the potential for biased results. In order to ensure high quality results, a discrete set of categories was defined for each variable (road type, weather, surface, visibility, and traffic) and a detailed description of each option was provided and explained to each video viewer. The video viewers were trained to leverage the capabilities of the developed Automated Visualization and Reduction Tool to review the video footage and annotate the environmental and traffic conditions for each trip.

3.3.1. Automated visualization and reduction tool

A NDS Automated Visualization and Reduction Tool was developed as part of this study with the capability of displaying two video files for the front and rear cameras as well as the time-series and radar data. The software synchronizes all three files for any given trip, which allows the user to consider multiple data sources simultaneously in the tool's graphical user interface (GUI). In addition to synchronizing video and time series data files for the purpose of manual video observation, the tool contains a feature that processes the video data and produces a visibility index that can be used to estimate visibility conditions that are difficult (and time consuming) to quantify in manual video observation. The GUI for the Visualization and Visibility Identification Tool is shown in Fig. 5.

Once the data sources are synchronized, valuable information can be derived quickly. The trip shown in Fig. 5 is an example of a trip that experienced both clear and adverse weather conditions. The continuous time series variables shown are: speed (red), acceleration (green), and headway (blue). The plots indicate that the driver reduced their speed by more than 20 km/hr at the onset of the heavy rain (shown in the front and rear cameras), and their speed fluctuated afterward. The radar map aggregated the provided radar data and shows objects detected in front of the subject vehicle. Finally, the image edge detection and histogram are used as part of the software's visibility estimation algorithm.

The visibility estimation algorithm applies methods that look for object boundaries and edges as a way to assess the existence of objects and their clarity in the image. This technique assumes that an image taken in adverse weather conditions is, generally, blurrier than the same image in clear weather. The front video footage is pre-processed by adjusting the video orientation and cropping unnecessary image components. The processing algorithm includes three levels of filters to smooth the image pixels and remove image noise: (1) Median Filter, (2) Gaussian Smoothing Low Pass Filter, and (3) Laplacian Filter. After the images are filtered, the image gradient (magnitude of the edge as calculated by the Laplacian filter) is computed as a percentage.

The visibility index (VI) is the output value assigned to an image to describe its visibility level. The visibility index is currently computed at 5 Hz, based on the available image processing speed. The VI is expressed as a percentage and classified in one of three levels: low, medium, or high, as shown in Eq. (1). The thresholds defining the three VI categories were ascertained using a training dataset as a guideline to correlate gradient values from videos with similar weather conditions.

$$VI = \begin{Bmatrix} Low & 0 \le \nabla < 0.56 \\ Med & 0.56 \le \nabla < 0.71 \\ High & 0.71 \le \nabla < 1 \end{Bmatrix} \tag{1}$$

where ∇ represents the computed image gradient.

The algorithm is heuristic; therefore, accurate results for all input images are not guaranteed, as the algorithm can get stuck in local minima. The accuracy of the visibility estimation algorithm depends on numerous factors including: the training data set, the filter magnitude interpreted, cutoff limits used for different weather conditions, and the input image quality. The Visibility Index (VI) is the resulting value given to an image to describe its visibility level. The VI is expressed as a percentage and classified in one of three levels: low, medium, or high.

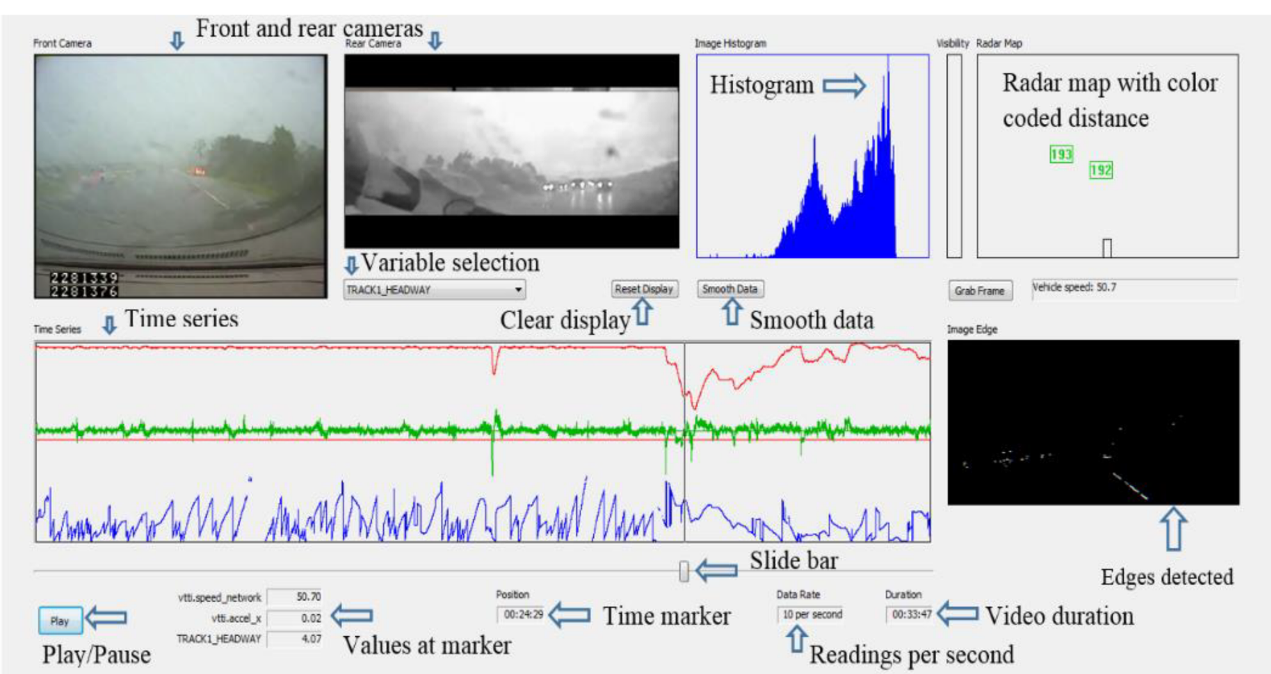


Fig. 5. Wyoming visualization and visibility identification tool.

More work is in progress to improve the methods for deriving the VI values, as well as to produce representative ranges of VI values for each classification (i.e., low, medium, high). As the visibility estimation algorithm is still under development and refinement, only experimental accuracies can be defined. Preliminary testing of the algorithm using 19 video files suggests a 79 percent accuracy (14 trips yielding results consistent with human observation; 2 trips yielding partially consistent results, and 3 trips yielding inconsistent results) (Ahmed et al., 2017, 2015).

3.4. Merge data files

Once the manual video observation is complete and conditions (i.e., weather, visibility, surface condition, road type, and traffic density) are reported for a group of trips, summary files describing statistical averages, standard deviations, coefficient of variations, and ranges of time series variables (including windshield wipers, speed, acceleration, braking, yaw rate, and headway), for each 1-min chunk can be generated automatically using the Analytic Tool. The driver demographic and survey data can be added to the data based on the random driver identifier provided, and geometric characteristics of the traversed trip route (extracted from the RID) can be added based on the first and last GPS coordinates recorded for each 1-min chunk. Once generated, these summary files can serve as the foundation for generating driving behavior models.

4. Results

Using the three complementary approaches, the research team received 11,164 potential weather-related trips and 22,328 matched trips in clear weather conditions (the same driver and route were collected, when available). In total, more than 11,000 h of driving were collected for adverse weather trips and their matching trips.

4.1. Driver demographics

The extracted dataset involved 1523 drivers between 16 and 99 years of age with the majority of the drivers in the young age group, 16 to 29 years old. Gender representation was balanced in most age groups, with the exception of a slight overrepresentation of female drivers between 20 and 24 years old. Fig. 6 compares the distribution from this study's NDS data sample with the entire SHRP2 database. In this representation, drivers are represented in the following age groups: young (below 24), middle (24–64), and old (above 64).

4.2. Verified weather-related trips

As indicated previously, the extraction process produced 11,164 potential weather-related trips that needed to be confirmed by means of the data reduction procedure. Due to the methodologies outlined in the three complementary trip extraction methodologies, a significant number of false-positives (trips flagged as potential weather-related trips but after data reduction were confirmed to have occurred in clear conditions) were expected. In total, 4094 freeway trips were verified to have adverse weather or roadway surface conditions, which is approximately 37% of the flagged trips. Preliminary data verification revealed that the research team received 3013 trips in rain; 234 trips in fog; 320 trips in snow; 317 trips in clear weather and wet surface condition; and 210 trips in clear weather and snow covered surface condition. Fig. 7 shows the number of received trips in each weather condition and each specific extraction method.

5. Conclusions

Initial acquisition of quality and representative data is imperative for conducting consequential research. Extraction of SHRP2 NDS trips in adverse weather conditions presented a unique challenge that required the development of creative and unique methods for leveraging the full extent of the SHRP2 data. Three complementary methodologies were developed to extract weather-related trips including wiper status, weather stations, and weather-related crashes. The first data acquisition method leveraged the front windshield wiper status, available directly from the NDS time series vehicle data. This method was successful in extracting trips with varying levels of precipitation; however, other weather and road surface conditions were not captured. Therefore, the second and third methods applied data gathered from the SHRP2 RID and NCDC to capture trips that passed through temporal-spatial domains with known adverse weather conditions.

The extracted NDS trips using the developed complementary approaches are representative (in driver age and gender) of the entire SHRP2 NDS database. After processing each extracted trip, 95% (method 1), 27% (method 2), and 42% (method 3) of the extracted trips were verified to have occurred on freeways in adverse weather conditions. These results indicate that the three complementary data acquisition methodologies are successful in identifying weather-related trips in the SHRP2 NDS database, and hold promise for providing solutions for similar research efforts leveraging other naturalistic driving studies worldwide. This unprecedented data source opens doors for developing innovative operational strategies that will improve the safety and mobility of the transportation network during adverse weather conditions.

6. Summary of lessons learned

1. Using only wiper status to capture weather conditions is not sufficient, as reported wiper settings are not uniform between different vehicles and is often missing from the SHRP2 NDS time series data.
2. Utilizing different weather sources (such as weather stations and weather-related crashes) increases the accuracy of assigning weather information to each NDS trip, while expanding capability of the research team to identify a variety of weather-related trips.
3. An efficient data reduction procedure is required because of the wide projection of an adverse weather event on an entire day; however, the utilization of this procedure enabled the collection of all trips impacted by an adverse weather event.

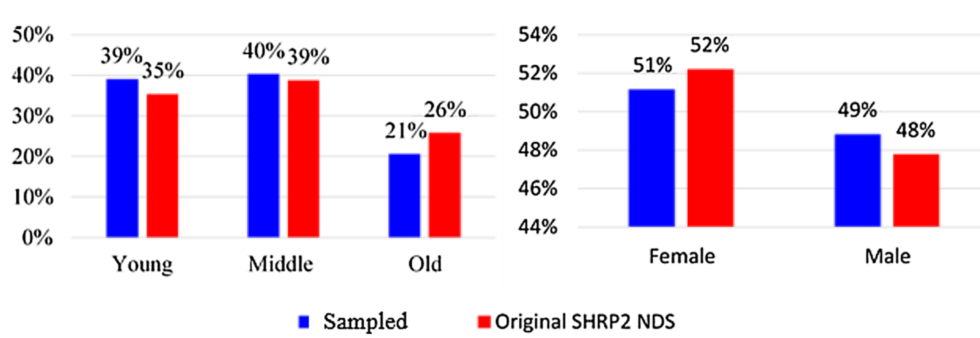


Fig. 6. Age and gender distribution for sampled data vs. Original SHRP2 NDS data.

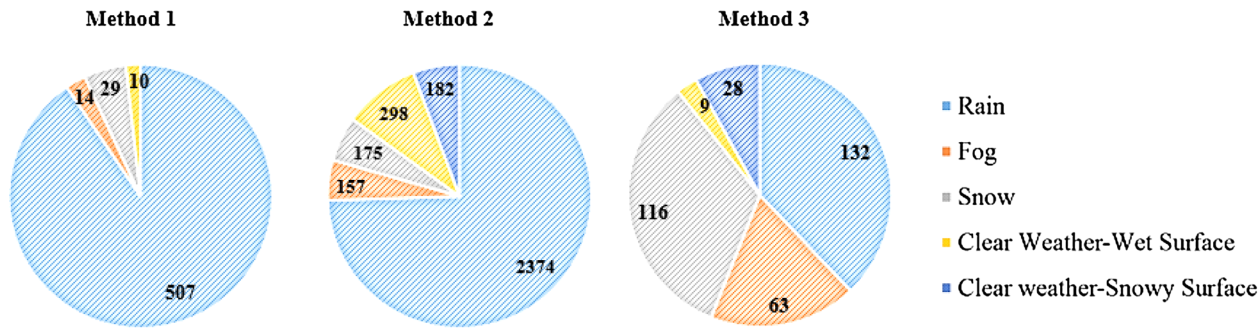


Fig. 7. Number of received trips in each weather condition for the three developed methodologies.

4. The procedure using 1-min chunks of time series data increased the capability of the research team to analyze the massive NDS dataset, while maintaining an adequate level of resolution to be used in future analyses.

Acknowledgements

This work was conducted through the second Strategic Highway Research Program (SHRP2), which is administrated by the Transportation Research Board (TRB) of the National Academies of Sciences, Engineering, and Medicine, and it was sponsored by the Federal Highway Administration (FHWA) in cooperation with the American Association of State Highway and Transportation Officials (AASHTO).

References

Ahmed, M., Abdel-Aty, M., Lee, J., Yu, R., 2014a. Real-time assessment of fog-related crashes using airport weather data: a feasibility analysis. Accid. Anal. Prev. 72, 309–317. https://doi.org/10.1016/j.aap.2014.07.004.
Ahmed, M., Abdel-Aty, M., Qi, S., Abuzwidah, M., 2014b. Synthesis of state-of-the-art in visibility detection systems' applications and research. J. Transp. Saf. Secur. 6, 183–206. https://doi.org/10.1080/19439962.2013.824055.
Ahmed, M., Abdel-Aty, M., Yu, R., 2012. Assessment of interaction of crash occurrence, mountainous freeway geometry, real-time weather, and traffic data. Transp. Res. Rec. J. Transp. Res. Board 51–59.
Ahmed, M.M., Ghasemzadeh, A., 2018. The impacts of heavy rain on speed and headway behaviors: an investigation using the SHRP2 naturalistic driving study data. Transp. Res. Part C Emerg. Technol. 91, 371–384. https://doi.org/10.1016/j.trc.2018.04.012.
Ahmed, M.M., Ghasemzadeh, A., 2017. Exploring the impacts of adverse weather conditions on speed and headway behaviors using the SHRP2 naturalistic driving study data. 96th Transportation Research Board Annual Meeting. Washington, D.C.
Ahmed, M.M., Ghasemzadeh, A., Eldeeb, H., Gaweesh, S., Clapp, J., Ksaibati, K., Young, R., 2015. Driver Performance and Behavior in Adverse Weather Conditions: An Investigation Using the SHRP2 Naturalistic Driving Study Data—Phase 1. DOI: 10.13140/RG.2.2.24061.05602 (RPRT). http://doi.org/10.13140/RG.2.2.24061.05602.
Ahmed, M.M., Young, R.K., Ghasemzadeh, A., Hammit, B., Elhashemi, A., Khan, N., Das, A., Eldeeb, H., 2017. Implementation of SHRP2 results within the Wyoming connected vehicle variable speed limit system: phase 2 early findings report and phase 3 proposal.
Andreescu, M.P., Frost, D.B., 1998. Weather and traffic accidents in Montreal. Canada. Clim. Res. 9, 225–230.
Andrey, J., Olley, R., 1990. The relationship between weather and road safety: past and future research directions. Climatol. Bull. 24, 123–127.
Brodsky, H., Hakkert, A.S., 1988. Risk of a road accident in rainy weather. Accid. Anal. Prev. 20, 161–176.
Das, Anik, Ghasemzadeh, Ali, Ahmed, Mohamed M., 2018. Analyzing the effect of fog weather conditions on driver lane-keeping performance using the SHRP2 naturalistic driving study data. J. Safety Res. 68, 71–80.
Eisenberg, D., Warner, K.E., 2005. Effects of snowfalls on motor vehicle collisions, injuries, and fatalities. Am. J. Public Health 95, 120–124.
El-tawab, S., Olariu, S., 2010. FIRMS : A Framework for Intelligent Road Monitoring System using Smart Sensors 1, 1–6.
Ghasemzadeh, A., Ahmed, M.M., 2017a. A probit-decision tree approach to analyze effects of adverse weather conditions on work zone crash severity using second strategic highway research program roadway information dataset. In: 96th Transportation Research Board Annual Meeting. Washington, D.C, pp. 1–17.
Ghasemzadeh, A., Ahmed, M.M., 2017b. A tree-based ordered probit approach to identify factors affecting work zone weather-related crashes severity in North Carolina using the highway safety information system dataset. 96th Transportation Research Board Annual Meeting. Washington, D.C.
Ghasemzadeh, A., Ahmed, M.M., 2017c. Drivers' lane-keeping ability in heavy rain: preliminary investigation using SHRP 2 naturalistic driving study data. Transp. Res. Rec. J. Transp. Res. Board 99–108. https://doi.org/10.3141/2663-13.
Ghasemzadeh, A., Ahmed, M.M., 2018a. Complementary parametric probit regression and nonparametric classification tree modeling approaches to analyze factors affecting severity of work zone weather-related crashes. J. Modern Transport. 1–12.
Ghasemzadeh, A., Ahmed, M.M., 2018b. Exploring factors contributing to injury severity at work zones considering adverse weather conditions. IATSS Res in press.
Ghasemzadeh, Ali, Ahmed, Mohamed M., 2018c. Utilizing naturalistic driving data for in-depth analysis of driver lane-keeping behavior in rain: non-parametric MARS and parametric logistic regression modeling approaches. Transport. Res. Part C: Emerg. Technol. 90, 379–392.
Ghasemzadeh, Ali, Ahmed, Mohamed M., 2018d. A framework to enhance the transferability of the SHRP2 naturalistic driving study by considering heterogeneity of driver behavior using spatial-temporal factors in a trajectory level. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE.
Ghasemzadeh, Ali, Hammit, Britton E., Ahmed, Mohamed M., Young, Rhonda Kae, 2018. Parametric ordinal logistic regression and non-parametric decision tree approaches for assessing the impact of weather conditions on driver speed selection using naturalistic driving data. Transport. Res. Rec 0361198118758035.
Hamilton, B.A., 2016. Ten-year averages from 2005 to 2014. Retrieved from Road Weather Management Program - NHTSA. http://www.ops.fhwa.dot.gov/weather/q1_roadimpact.htm.
Hammit, Britton E., et al., 2018. Evaluation of weather-related freeway car-following behavior using the SHRP2 naturalistic driving study database. Transport. Res. Part F: Traffic Psychol. Behav. 59, 244–259.
Hankey, J.M., Perez, M.A., McClafferty, J.A., 2016. Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets (RPRT). Tech Transportation Institute, Virginia.
Highway Capacity Manual, 2016. Transp. Res. Board. Natl. Res. Counc., Washington, DC.
Kamrani, Mohsen, Arvin, Ramin, Khattak, Asad J., 2018a. Extracting useful information from Basic Safety Message Data: an empirical study of driving volatility measures and crash frequency at intersections. Transport. Res. Rec 0361198118773869.
Kamrani, Mohsen, Arvin, Ramin, Khattak, Asad J., 2018b. Safety impacts of automated vehicles in mixed traffic. Transportation Research Board 97th Annual Meeting, Washington DC, United States.
Kamrani, Mohsen, Arvin, Ramin, Khattak, Asad J., 2018c. Analyzing highly volatile driving trips taken by alternative fuel vehicles. Transportation Research Board 97th Annual Meeting, Washington DC, United States.
Khan, Md Nasim, Ghasemzadeh, Ali, Ahmed, Mohamed M., 2017. Investigating the impact of fog on freeway speed selection using the SHRP2 naturalistic driving study data. Transport. Res. Rec 0361198118774748.
Knapp, K.K., Smithson, L.D., Khattak, A.J., 2000. Mobility and Safety Impacts of Winter Storm Events in a Freeway Environment (RPRT). Center for Transportation Research and Education, Iowa State University.
Kyte, M., Khatib, Z., Shannon, P., Kitchener, F., 2001. Effect of weather on free-flow speed. Transp. Res. Rec. J. Transp. Res. board 60–68.
Maze, T., Agarwai, M., Burchett, G., 2006. Whether weather matters to traffic demand, traffic safety, and traffic operations and flow. Transp. Res. Rec. J. Transp. Res. board 170–176.
Perez, M., Mclaughlin, S., Kondo, T., Antin, J., Mcclafferty, J., Lee, S., Hankey, J., Dingus, T., 2016. Transportation safety meets big data: the SHRP 2 naturalistic driving database. J. Soc. Instrum. Control Eng. 55, 415–421.
Smadi, O., Hawkins, N., Hans, Z., Bektas, B.A., Knickerbocker, S., Nlenanya, I., Souleyrette, R., Hallmark, S., 2014. Naturalistic Driving Study: Development of the Roadway Information Database, SHRP 2 Report S2-S04A-RW-1 (RPRT).