Journal of Hazardous Materials 167 (2009) 374–382

Contents lists available at ScienceDirect

Journal of Hazardous Materials

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

Utilization of accident databases and fuzzy sets to estimate frequency of HazMat transport accidents

Yuanhua Qiao [a], Nir Keren [b,*], M. Sam Mannan [a]

[a] Mary Kay O'Connor Process Safety Center, Chemical Engineering Department, Texas A&M University, College Station, TX 77843-3122, USA
[b] Department of Agricultural and Biosystems Engineering [ABE] Graduate Faculty, Human Computer Interaction [HCI] Iowa State University 102 I Ed Building II Ames, IA 50011-3130, USA

ARTICLE INFO

Article history:
Received 6 June 2008
Received in revised form
21 November 2008
Accepted 1 January 2009
Available online 5 February 2009

Keywords:
Transportation risk analysis
Hazardous materials transportation
Accident frequency
Negative binomial regression
Fuzzy logic model

ABSTRACT

Risk assessment and management of transportation of hazardous materials (HazMat) require the estimation of accident frequency. This paper presents a methodology to estimate hazardous materials transportation accident frequency by utilizing publicly available databases and expert knowledge. The estimation process addresses route-dependent and route-independent variables. Negative binomial regression is applied to an analysis of the Department of Public Safety (DPS) accident database to derive basic accident frequency as a function of route-dependent variables, while the effects of route-independent variables are modeled by fuzzy logic. The integrated methodology provides the basis for an overall transportation risk analysis, which can be used later to develop a decision support system.

1. Introduction

More than 3.1 billion tons of hazardous materials (HazMat) are shipped annually in the United States [1]. According to Department of Transportation (DOT) statistics, 156,442 HazMat transportation accidents occurred from 1995 to 2004, resulting in a total of 221 fatalities and 3143 injuries [2]. The public, along with agencies such as the DOT and the Federal Emergency Management Administration, show an increasing concern with the risks associated with HazMat transportation. Few regulations and rules have been set to regulate HazMat transportation activities. In fact, existing regulations mainly address hardware and procedures, and compliance with those regulations does not necessarily guarantee the desired reduction in the level of risk. Notably, a significant reduction may be gained by selecting the route with relatively less risk. Selection of the best route for HazMat transportation involves comparing alternatives in the domain of risk. This optimization approach is possible as long as risk can be quantified.

Risk is a combination of two parameters: frequency and the magnitude of the consequence; thus, accident frequency estimation is essential for risk analysis. Currently, the most popular data cited for accident frequency takes only a few factors into consideration. This paper presents a methodology to estimate the accident frequency for different types of roads by incorporating the effects of a larger number of parameters, including the nature of truck configurations, operating conditions, environmental factors, and road conditions.

2. Background

2.1. Accident frequency assessment

Accident frequency can be defined as the number of accidents per unit of road (mile, kilometer, etc.). The frequency can be computed by dividing the number of accidents by the number of vehicle miles, which is the corresponding exposure measure of opportunities for an accident to occur.

There are three basic options to assess accident frequency with reasonable accuracy. The first is to obtain at least one database and analyze both accident data and travel data for the specific conditions under investigation (assuming that the dataset is structured to support distinctions between the desired variables). The second option is to access state databases for specific routes. Frequently, states have accident data and travel data for major state highways. A third option is to use an existing limited analysis of databases and apply the results to a specific route of interest. All three options are harnessed in this work.

Detailed analyses of several publicly available databases [3–7] have made it possible to specify accident frequency on a per-mile basis. One of the most detailed analyses of such data was con-

* Corresponding author. Tel.: +1 515 294 2580 (ABE)/5685 (HCI); fax: +1 515 294 1123.
E-mail address: nir@iastate.edu (N. Keren).

0304-3894/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhazmat.2009.01.097

ducted by Harwood and Russell [5]. On the basis of computerized data files from three states – California, Illinois, and Michigan – Harwood and Russell calculated accident frequency by using the number of reported accidents and the total number of truck-miles traveled. Accident frequency was assessed as a function of road type, truck type, and population density. In the report for Argonne National Laboratory [6], Harwood and Russell's statistics were divided into two road categories: interstates and non-interstates (state highways), and into three population density categories: urban, suburban, and rural. Bubbico et al. [8,9] pointed out that both route-independent and route-dependent parameters affect risk, but their work did not yield a methodology for estimating accident frequency. The results from Argonne National Laboratory also failed to improve the frequency data sensitivity, in spite of incorporating more parameters describing the nature of the roads, characteristics of the trucks, environmental factors, and driver conditions.

This paper presents an integrated methodology to estimate accident frequency by incorporating the effects of various parameters, including both route-dependent and route-independent variables.

2.2. Fuzzy logic model

Conventionally, a mathematical model of a system is constructed by analyzing input–output measurements from the system. However, an additional important source of information about engineering systems is human expert knowledge, known as linguistic information. It provides qualitative instructions and descriptions of the system. While a conventional mathematical model fails to include this type of information, a fuzzy model can conveniently incorporate it.

The core technique of fuzzy logic is based on three basic concepts: (1) fuzzy set: unlike crisp sets, a fuzzy set has a smooth boundary, i.e., the elements of the fuzzy set can be partly within the set. Membership functions are employed to provide gradual transition from regions completely outside a set to regions completely in the set; (2) linguistic variables: variables that are qualitatively, as well as quantitatively, described by a fuzzy set. Similar to a conventional set, a fuzzy set can describe the value of a variable; (3) fuzzy "if-then" rules: a scheme, describing a functional mapping or a logic formula that generalizes an implication of two-valued logic. The main feature of the application of fuzzy "if-then" rules is its capability to perform inference under partial matching. It computes the degree the input data matches the condition of a rule. This matching degree is combined with the consequence of the rule to form a conclusion inferred by the fuzzy rule.

A fuzzy "if-then" rule associates a given condition to a conclusion, using linguistic variables and fuzzy sets. The most common fuzzy model, the Mamdani model [10], consists of the following fuzzy "if-then" rules that describe a mapping from $U_1 \times U_2 \times \cdots \times U_r$ to W:

$$R_i: \text{ If } x_1 \text{ is } A_{i1}, \ldots, \text{ and } x_r \text{ is } A_{ir}, \text{ then } y \text{ is } C_i \quad (1)$$

where x_j ($j = 1, 2, \ldots, r$): input variables; y: output variable; A_{ij}: fuzzy sets for x_j; C_i: fuzzy sets for y.

The relationship between the various parameters and accident frequency is difficult to express by a function; however, it is possible to express the relationship among parameters using the fuzzy if-then rules. For example, if a driver is not experienced, then accident frequency is high. This type of association can be conveniently incorporated into fuzzy models. This characteristic is especially important, given the complexity of transportation conditions and the level of human experience/knowledge about the system.

Fuzzy logic is a form of multi-valued logic derived from fuzzy set theory to allow reasoning that is approximate rather than precise. Fuzzy logic is no less precise than any other form of logic; it is an organized and mathematically sound method of handling inherently imprecise concepts. Fuzzy if-then rules are selected based on previous findings or based on experts' experience. Findings from previous studies will be utilized for setting up the rules, if the related parameters have been estimated before. For example, as detailed in Section 2.1, the effects of parameters, including road type and population density have been previously studied. It is important to emphasize that when knowledge from experts is incorporated, this knowledge may introduce subjectivity into the categories used; however, the resulting values from the fuzzy sets will be absolutely accurate. For example, assume that there is a need to categorize HazMat tanks into hazardous categories. A group of experts may define a 500 gallon tank to be low-level hazardous, a 1000 gallon tank to be moderately hazardous, and 5000 gallon tank to be highly hazardous. Using fuzzy sets, one can determine the level of membership of a 3500 gallon tank in the highly hazardous category and its level of membership in the moderately hazardous category. While one can argue that experts' labeling categories for tank volumes are subjective, the level of membership of a given tank volume in any one of the groups, as determined by fuzzy logic, is not questionable. Further explanations on subjectivity control efforts are available later. Conclusions from earlier studies are employed when the rules related to those parameters are set up. If there are no previous studies available, human intelligence, scientific knowledge and working experience will be applied to derive the rule set. Since fuzzy logic deals with reasoning that is approximate rather than precise, there could be some subjectivity involved in the approach; however, it is no less precise than any other form of logic, it applies the best available information and it will become more precise with the increase of our understanding of the problem.

3. Methodology

3.1. Data and database analysis

Since accident frequency can be computed by dividing the number of accidents by the number of vehicle miles, both accident and the corresponding exposure measure of vehicle-mile data are needed to assess accident frequency.

The Hazardous Material Information System (HMIS) is the national database of HazMat transportation accidents, encompassing container types, consequences of the accidents, and other information. However, accidents occurring on intrastate roads and accidents not resulting in spill are not recorded in this database. Battelle [7], in a report to the Federal Motor Carrier Safety Administration, suggested supplementing HMIS with additional databases that consist of data on non-spill accidents and other spill accidents (especially intrastate accidents).

The Department of Public Safety (DPS) accident databases from each state consist of accident data gathered from state highways. These databases can be easily grouped based on the type of road, which makes it relatively easy to assess the accident frequency for a specific road. Some parameters that affect accident frequency, specifically road and environmental conditions, are available in DPS databases. Thus, DPS databases can be utilized to establish a model for estimation of accident frequency by incorporating the effects of those route-dependent parameters available in the datasets.

In addition to the number of accidents, the number of miles traveled (exposure data) is needed. The most commonly cited exposure data source is the Commodity Flow Survey (CFS) [11], which is generated from a 5-year economic census, and was last conducted in 2002. It provides information on commodities shipped, their value, weight, and mode of transportation. Exposure data on state highways can also be obtained from state DOT's or transportation institutes. In many cases, both the data from CFS and data from state

Table 1
DPS accident database – sample of records of accidents occurred in US290.

COUNTY district	MILE1 milepoint	WEATHER	SURF_CON	ROAD_CON	ROADWAY	INTERSECT
101	315	1	1	0	1	4
101	356	2	2	0	1	3
101	313	1	1	0	1	2
101	332	1	1	0	1	4
101	381	2	4	5	3	4
101	364	8	4	5	3	4
101	384	8	4	5	1	4
101	308	8	4	5	3	4
101	243	8	4	5	1	4
101	211	3	1	4	9	3
101	367	2	2	2	9	3

COUNTY district: indicate county number derived from DPS County Listing.
MILE1 milepoint: control/section number. For tool way accidents, the station number is recorded. For all accidents on highways, record the control and section number as coded from the District Control-Section Maps. Accidents not on numbered highways, code the first five characters of the street or county road name.
WEATHER: 1, clear (cloudy); 2, raining; 3, snowing; 4, Fog; 5, blowing dust; 6, smoke; 7, other; 8, sleeting.
SURF_CON: surface condition, 1, dry; 2, wet; 3, muddy; 4, snowy/icy.
ROAD_CON: road condition, 0, no defects; 1, holes; ruts, etc., 2, defective shoulders; 3, foreign material on surface; 4, high water or flood debris; 5, slick surface; 6, obstruction in road not lighted (night); 7, obstruction in road not marked (day); 8, 9, road under construction.
ROADWAY: roadway related, 1, on roadway; 2, off roadway on shoulder; 3, off roadway beyond shoulder.
INTERSECT: intersection related, 1, intersection; 2, intersection related; 3, driveway access; 4, non intersection.

DOT's or transportation institutes are needed to assess the exposure data for specific roads.

3.2. Parameter analysis

Accident frequency is affected by a large number of parameters such as the nature of the roads, characteristics of the trucks, environmental factors, and driver conditions. Some specific risk reduction measures that may have been implemented should also be considered in assessing accident frequency. Previous research has shown that considerations, such as urban versus rural and divided versus undivided highway, have a direct influence on accident frequency [5]. It is also been proposed that location-specific conditions, such as vehicle speed limit, topographical conditions, excessive grade, obstructions to vision, poorly designed intersections, etc., can have a significant localized influence on accident frequency. Weather conditions, such as rain, fog, storms, icing, wind, or tornado conditions, can also influence accident frequency. Similarly, driver training programs, fleet maintenance, speed monitoring, driver stress level, driver drinking-habits, and other characteristics specific to an individual carrier can influence accident frequency [4].

Based on the variables available in the databases (see Section 3.1), the following two groups of parameters are considered in this study: route-dependent parameters (available in DPS databases) and route-independent parameters (not from DPS databases).

The route-dependent parameters considered in this study include:

- Lane number (x_1)
- Weather (x_2)
- Population density (x_3)

Several route-independent parameters that affect accident frequency are as follows:

- Truck configuration (y_1)
- Container capacity (y_2)
- Driver experience (y_3)

The parameters were selected to represent the effects of various conditions including road, truck, environment, and carrier related conditions.

Since the affecting parameters are available from more than a single source, frequency assessment efforts need to incorporate data from different sets. This has been accomplished as follows: basic accident frequencies based on route-dependent parameters are derived from the DPS database. Then, the effects of route-independent parameters on accident frequencies are derivable from other databases such as HMIS, or by incorporating expert knowledge. It is essential to use fuzzy logic in the proposed methodology for the following two reasons: (1) the information available in HMIS is on nationwide transportation activities, while the data derived from DPS is for specific roads, i.e., data obtained from HMIS cannot be applied directly; (2) the effects of several of the parameters on the frequency cannot be derived from any database; thus, expert judgment must be employed in the assessment.

To summarize, the procedure to estimate accident frequency is as follows:

(1) Number of accidents is derived from the DPS databases as a function of route-dependent parameters.
(2) The corresponding vehicle-mile data are obtained from state DOT's or transportation institutes and from the 2002 CFS. The basic accident frequency is obtained by dividing the number of accidents by the number of miles traveled.
(3) The basic accident frequency is modified by considering the effects of route-independent parameters. Fuzzy logic is employed to incorporate expert knowledge. The membership functions of these parameters are built based on the data available in the HMIS database or based on expert experience.

The flow chart in Fig. 1 presents the process of utilizing the variety of data sources to establish an accident frequency assessor.

Following the development of the methodology, a case study assessment of HazMat accident frequency on a section from Texas highway US 290, will be presented.

3.3. Basic accident frequency assessment

Table 1 consists of data extracted from the Texas DPS database, which contains records on accidents that occurred on US 290 in 1999. A large number of parameters are collected to record information on the conditions of each accident (see first row of Table 1). In this study, we incorporate the effects of three parameters from DPS in the model. In the DPS databases, most of the parameters are treated as linguistic variables, and the numeric values in the table

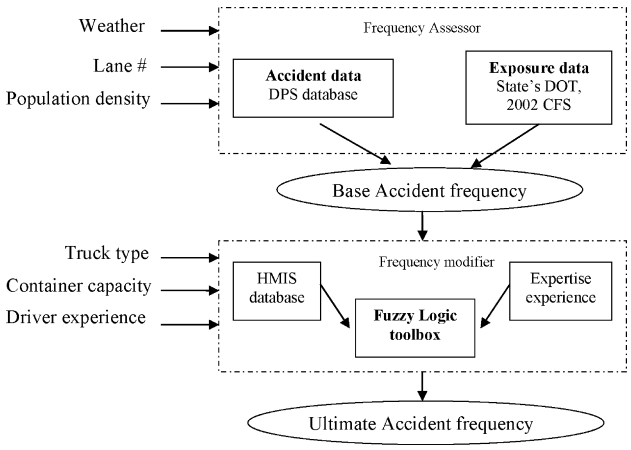

Fig. 1. Model for estimation of accident frequency.

represent the values of those linguistic parameters. For example, the variable "*surface condition*" has four linguistic values: 1, dry; 2, wet; 3, muddy; and 4, snowy/icy. A brief definition of each entry level for the parameters is presented at the bottom of Table 1.

DPS data for 10 years (from 1992 to 2001) is used in this study. The data was sorted by using Matlab to first obtain the number of accidents under all the given conditions in the databases. Then the number of accidents under any condition was estimated as a function of the three route-dependent parameters. An appropriate count data model was needed for the estimation. Four analytical approaches were considered for this purpose: linear regression, Poisson regression, negative binomial model, and Bayesian estimation. The following paragraphs discuss these approaches.

- Early works in empirical analysis used multiple linear regression models. However, these models suffer from several methodological limitations and practical inconsistencies. The two major deficiencies are: (a) linear regression assumes a normal distribution of the dependent variable, which is not valid for count (accident) data, and (b) linear regression may produce negative estimates for the dependent variable.
- The Poisson model has several advantages in comparison to the normal regression model. It assumes that the data follows a Poisson distribution, a distribution frequently encountered when events are counted. Despite its advantages, Poisson regression assumes that the variance and mean of the dependent variable are equal. However, it is quite common for the variance of data to be substantially higher than the mean. This phenomenon is known as "over-dispersion." Over-dispersion leads to invalid *t*-tests of the estimated parameters.
- The negative binomial regression model has the same advantage as the Poisson model, in that it assumes a distribution frequently encountered when events are counted. At the same time, it does not have the restriction that the variance and mean of the dependent variable have to be equal: it allows the variance of the dependent variable to be larger than the mean.
- Bayesian estimation can combine sample information with other information that may be available prior to collecting the sample. In a Bayesian model, each input independent variable has a probability distribution that is a function of one parameter (known as prior). This approach is useful when uncertainty exists in input variables. However, in DPS databases, the four affecting parameters have been defined and categorized.

Therefore, the negative binomial regression model is the best choice to assess the number of accidents. The regression model is derived from the statistical software SAS, thereby estimating the number of accidents under any condition.

Exposure data was obtained from state DOT or transportation institutes and the 2002 CFS. Since the number of accidents is a function of route-dependent parameters, the corresponding number of miles traveled needs to consider the effects of the same parameters, i.e., the exposure data must be disaggregated by the same factors.

Finally, the basic accident frequency was obtained by dividing the number of accidents by the number of miles traveled. This frequency is expressed as $f_{\text{basic}}(x_1,x_2,x_3)$. Fig. 2 is a schematic representation of the algorithm of estimation of basic accident frequency.

3.4. Modification to the basic accident frequency data

Basic frequency data needed to be modified to incorporate the effects of the following route-independent parameters: *truck configuration* (y_1), *container capacity* (y_2), and *driver experience* (y_3). These three parameters are not road-related variables, and their effects on the frequency are independent from road conditions. In this study, fuzzy Mamdani models [12] were employed to assess the effects of y_1, y_2, and y_3 on the frequency. A modifier, expressed as m_i (i = 1–3), was generated for each of these three parameters; the ultimate accident frequency is expressed as:

$$f_{\text{ultimate}} = f_{\text{basic}} \times (m_1 \times m_2 \times m_3)$$

y_1, y_2, and y_3 are treated as linguistic variables. The frequency modifiers are viewed as linguistic variables as well. Each linguis-

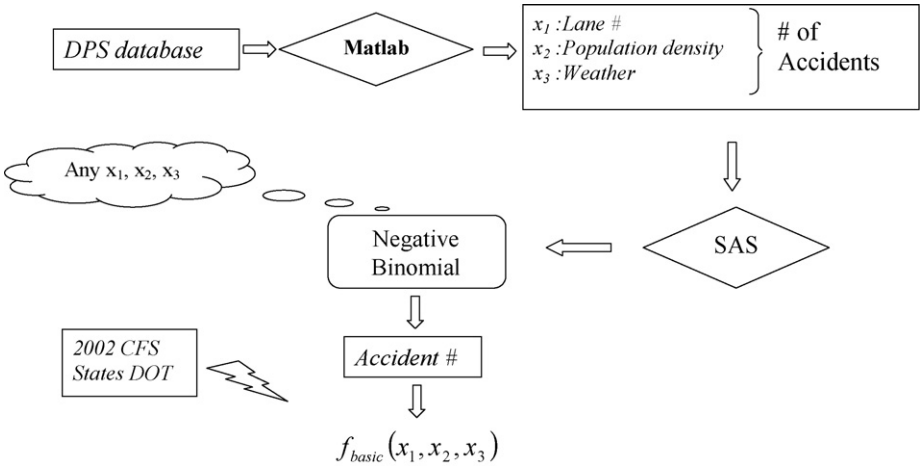

Fig. 2. Schematic representation of an algorithm of estimating basic accident frequency.

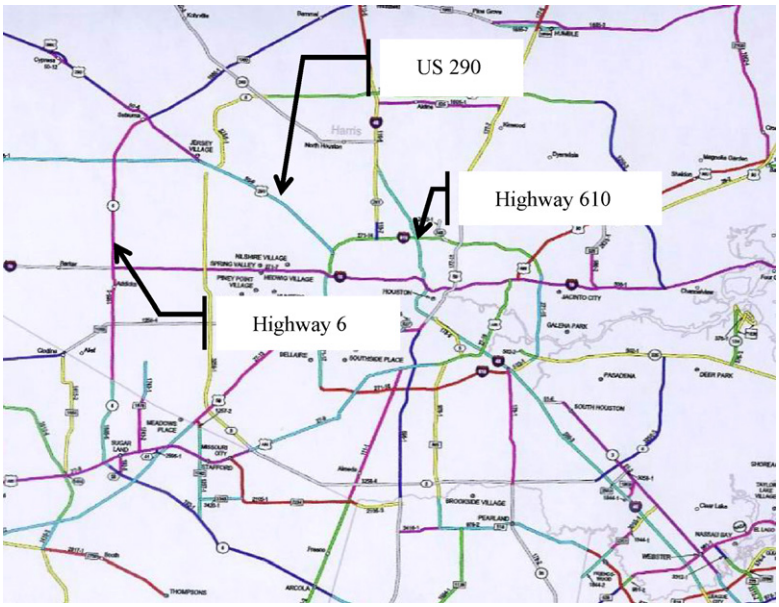

Fig. 3. Control section map for Houston TX.

tic variable was defined by several fuzzy sets. The membership function for each fuzzy set was determined either by expert experience or from available data. Analysis was performed on the HMIS database to develop the membership functions for *truck configuration*. The membership functions for *driver experience* and *container capacity* were determined based on expert experience.

There is as yet no fixed, unique, and universal rule or criterion for selecting a membership function for a particular fuzzy set. A good membership function is determined by the user based on his/her scientific knowledge, working experience, and recognition of the actual need for the particular application in question.

To increase membership functions' quality, when expert knowledge is utilized, it is necessary to apply the following two procedures: (1) carefully select the experts; and, (2) control for subjectivity.

Experts selected for the procedure in this study are nationally and internationally recognized loss control experts, each one having more than 20 years experience in Quantitative Risk Analysis with emphasis on Probabilistic Risk Assessment. To control for subjectivity, group decision making procedures such as Delphi [16] or Analytical Hierarchy Process [17] are needed to aggregate the data from experts. In this study we utilized Delphi.

When input functions were established, the fuzzy "if-then" rules were built to associate each affecting variable, y_i, to the corresponding modifier, m_i. For example, *driver experience* (y_3), as shown in Fig. 4, is expressed by three fuzzy sets: novice, medium, and experienced. The *driver experience modifier* variable (m_4), as shown in Fig. 5, also includes three fuzzy sets: low, medium, and high.

The "if-then" rules were set up as:

If *driver experience* is novice, then the *driver experience modifier* is high.
If *driver experience* is medium, then the *driver experience modifier* is medium.
If *driver experience* is experienced, then the *driver experience modifier* is low.

After setting-up the rules, defuzzification needs to be performed to obtain a numerical output value. Defuzzification is the process of producing a quantifiable result in fuzzy logic. A fuzzy system has a number of rules that transform a number of variables into a "fuzzy" result, that is, the result is described in terms of membership in fuzzy sets.

A useful defuzzification technique must first combine the results from the rules. The most typical fuzzy set membership function has the graph of a triangle. If this triangle were to be cut in a straight horizontal line somewhere between the top and the bottom, and the top portion were to be removed, the remaining portion forms a trapezoid. Typically, the first step of defuzzification is chopping off parts of the triangle to form trapezoids (or other shapes if the initial shapes were not triangles). For example, if the output has "low (15%)", then the area of the triangle from 15% and up will be removed. In the most common technique, the trapezoids from all input functions are then superimposed one upon the other, forming a single geometric shape. Then, the centroid of this shape, called the fuzzy centroid, is calculated. The x coordinate of the centroid is the defuzzified value.

FIS editor in Matlab was used in this study to defuzzify the process based on input data in order to derive the output modifier. The value obtained after the defuzzification on the driver experience parameter was the *driver experience modifier* (m_3) as shown in Fig. 10. Other modifiers were obtained similarly.

4. Case study

As mentioned earlier, this case study addresses sections from US 290 in Texas, a road connecting Houston and Austin. Fig. 3 presents the control section map for the Houston area. The sections under study are between Highway 610 and Highway 6.

4.1. Calculation of number of accidents by SAS

As previously noted, DPS datasets from 1992 to 2001 were employed in this study, with 9536 accidents recorded for this period. The data were sorted by Matlab and input into SAS.

4.1.1. Models estimation

As described in Section 3.3, the negative binomial regression model was employed to assess the number of accidents. Eq. (2) is the standard format for negative binomial regression. The number of accidents under any conditions (denoted as POP, NUMN_LN, and

Table 2
Parameter estimation.

Parameter	Parameter categories	Parameter estimates
$\beta_0 = 3.1703$		
POP (β_1)	0	−1.5727
	1	−6.0371
	3	−1.7244
	4	−2.2201
	9	0
NUMN_LN (β_2)	4	2.3151
	5	−2.359
	6	2.6869
	8	2.5499
	10	0
WEATHER (β_3)	1	1.9717
	2	0

Table 3
Parameter meaning in DPS.

Parameter	Parameter categories	Parameter meaning
POP (β_1)	0	Rural
	1	Towns under 2499 population
	3	2500–4999 population
	4	5000–9999 population
	9	250,000 population and over
NUMN_LN (β_2)	4	Number of lanes = 4
	5	Number of lanes = 5
	6	Number of lanes = 6
	8	Number of lanes = 8
	10	Number of lanes = 10
WEATHER (β_3)	1	Clear (cloudy)
	2	Raining (other)

WEATHER) was estimated using this equation:

$$N(i,j,k) = e^{(\beta_0+\beta_{1i}+\beta_{2j}+\beta_{3k})} \quad (2)$$

where N: number of accidents, i, j, k: notations for population density, number of lanes, weather condition, β_0: intercept of regression equation, β_1: regression coefficient for variable *population density*, β_2: regression coefficient for variable *number of lanes*, β_3: regression coefficient for variable *weather condition*.

The estimated negative binomial regression coefficients for the model are shown in Table 2. The definitions of each value of the parameters of DPS are given in Table 3.

Negative binomial distribution is a discrete probability distribution. It can be used to describe the distribution arising from an experiment consisting of a sequence of independent trials, subject to several constraints. In a series of independent Bernoulli trials, with constant probability p of a success, let the random variable X denote the number of trials until r successes occur. Then X has a negative binomial distribution with parameters p and $r = 1, 2, 3, \ldots$, and

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r$$

for $x = r, r+1, r+2, \ldots$.

The dependent variable (number of accidents) is a count variable, and the regression models the log of the expected count as a linear function of the predictor variables, which include population density, number of lanes, and weather condition. We can interpret each regression coefficient as follows: for 1 unit change in the predictor variable, the difference in the logs of expected counts of the response variable is expected to change by the respective regression coefficient, given the other predictor variables in the model are held constant.β_0, Intercept: This is the negative binomial regression estimate when all variables in the model are evaluated at zero. In this study, when the population density, number of lanes, and weather condition are zero, the log of the expected count for accidents is 3.1703 units.

$\beta_1, \beta_2, \beta_3$, Coefficients: This is the negative binomial regression estimate for 1 unit increase in the respective parameter categories, given the other variables are held constant in the model. For example, if the weather category were to increase 1 unit at category 0, the difference in the logs of expected accident counts would be expected to increase by 1.9717 units, while holding the other variables in the model constant.

4.1.2. Over dispersion and goodness of fit

While Poisson distribution is preferred for the suggested model, it was rejected due to over dispersion [13]. Negative binomial distribution is not restricted by over dispersion. Negative binomial regression was accepted following a Pearson's chi-square to test for goodness of fit ($\alpha = 0.05$).

4.2. Exposure data assessment

Exposure data for US 290 was obtained from the Texas Transportation Institute (TTI). The exposure data was disaggregated by the number of lanes and population density group. The disaggregated data was then segmented further to incorporate the effects of weather.

As stated in Section 3.2, weather conditions such as rain, fog, storms, icing, wind, or tornado conditions can influence accident frequency. With respect to the location of the route, the rain condition will affect the accident frequency more than the others, since it occurs much often than fog, storms, icing, or tornado in Texas. Wind conditions have much less impact on accident frequency than do rain conditions. Therefore, only rain conditions are considered as a weather impact factor. According to the National Climate Data Center [14], the normal annual precipitation in this area is 47.84 in.

Table 4
Exposure (vehicle miles) data for control sections 58 and 59 on US290.

Exposure data (vehicle miles)			NUM_LN				
			4	5	6	8	10
Weather							
1	POP	0	498870024	1405378	951118019	–	3931089
		1	–	–	4152767	–	–
		3	444307187	–	14353884654	–	–
		4	243085504	–	213621596	–	–
		9	54087951	–	2601401642	1447712804	249977292
2	POP	0	26256317	73967	50058843	–	206899
		1	–	–	218566	–	–
		3	23384589	–	755467613	–	–
		4	12793974	–	11243242	–	–
		9	2846734	–	136915876	76195410	13156700

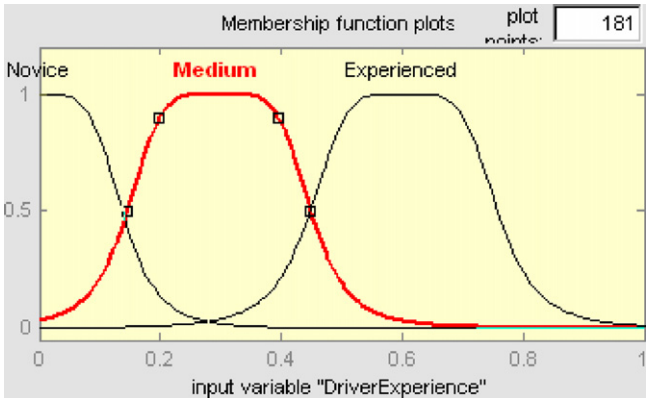

Fig. 4. Driver experience membership function.

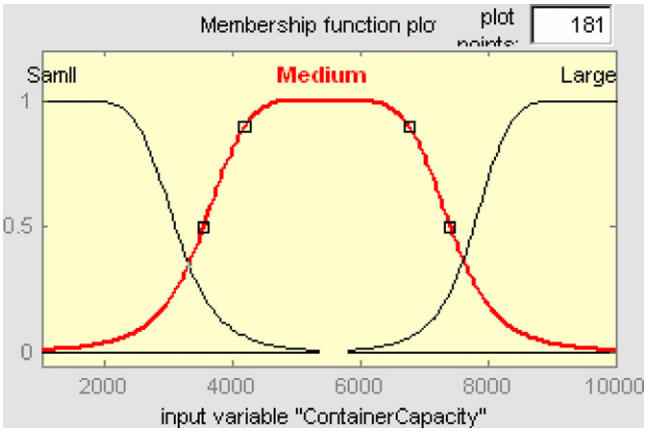

Fig. 6. Container capacity membership function.

The rain rate was modeled by Crane [15], whose study shows that the percentage of raining time over a year in the Houston area is about 3.9%. The annual precipitation assessed by multiplying the raining rate by the raining time is about 53.3 in., very close to the data reported by National Climate Data Center. Thus, the assessment by Crane [15] was employed here, but was approximated as 5% of the raining percentage over a year. We assumed that weather conditions would affect the corresponding vehicle mile data equally over the number of lanes and population groups.

The completely disaggregated exposure data are shown in Table 4. By dividing the estimated number of accidents (calculated in section 4.1) by the corresponding exposure data (listed in Table 4), the basic accident frequency was calculated.

4.3. Fuzzy logic modification

The basic frequency data was modified to incorporate the effects of *truck configuration* (y_1), *container capacity* (y_2), and *driver experience* (y_3), which were treated as linguistic variables. The frequency modifiers were viewed as linguistic variables as well. Each linguistic variable was defined by several fuzzy sets, and the membership function for each fuzzy set was determined either by experts or from data, as described earlier.

The membership function plots and the corresponding modifier plots for *driver experience* and *container capacity* are shown in Figs. 4–7.

A fuzzy set description of *driver experience* is given in Fig. 4, which captures the essence of the gradations between experience ranges. Three fuzzy sets represent the different experience groups, ;novice, medium, and experienced. Each membership function in the figure is represented by a curve that indicates the assignment of a degree of membership in a fuzzy set to each variable within the domain of the variable involved – *driver experience*. When driver experience is zero, the driver is almost completely a novice. The fuzzy set describing the driver experience modifier (presented in Fig. 5), includes three fuzzy sets representing different degrees of impact of *driver experience* on accident frequency. These degrees are low, medium, and high. The fuzzy sets in Fig. 5 have been established similarly to the sets in Fig. 4.

Three fuzzy sets are assigned for *container capacity* for the following memberships: small, medium, and large. The membership functions for the three curves are assigned symmetrically along the domain of the container capacity (gallons), as shown in Fig. 6. If the container capacity is 10,000 gallons or higher, the degree of membership in the large fuzzy set is equal to 1. Fig. 7 presents the fuzzy sets for the capacity modifier. Similar to Fig. 6, three variables or fuzzy sets are assigned for the capacity modifier: low, medium, and high. It is important to mention that even though the total number of the fuzzy sets for the input parameter *container capacity* is equal to that for the output parameter capacity modifier in this study, it is not necessary that the number of fuzzy sets for input and output parameters be equal for fuzzy logic analysis. The fuzzy "if-then" rules will relate all the input fuzzy sets and output fuzzy sets, regardless of the number of input or output fuzzy sets.

The four fuzzy sets in Fig. 8 are for *truck configuration* parameters. These parameters are passenger, single unit, single trailer, and double trailer. Harwood and Russell [5] studied the effects of truck configuration on accident frequency. Their study shows that accident frequency can increase by almost 50% if the truck configuration

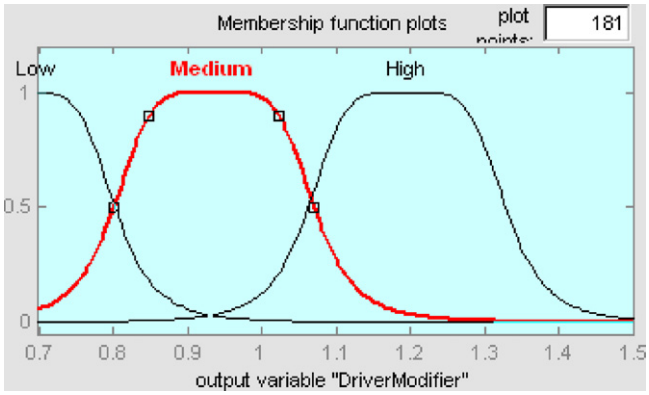

Fig. 5. Driver modifier membership function.

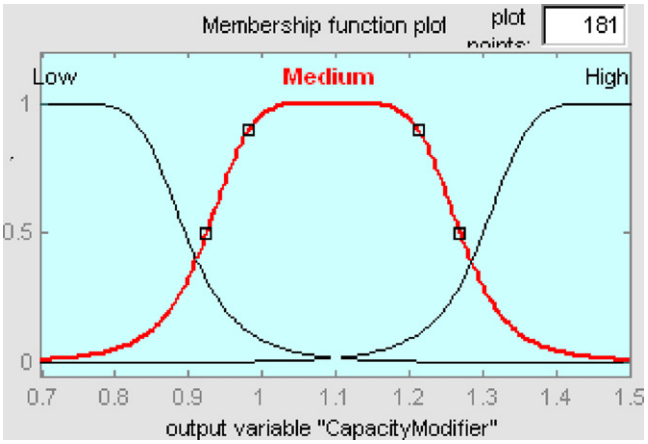

Fig. 7. Container modifier membership function.

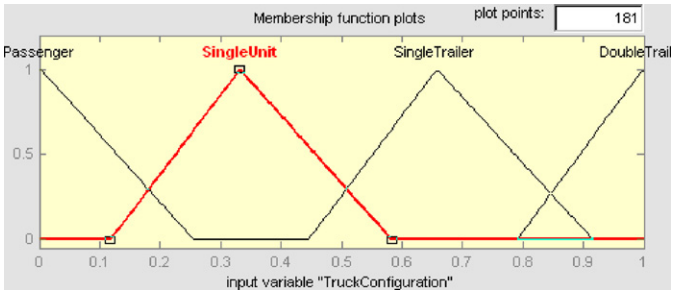

Fig. 8. Truck configuration membership function.

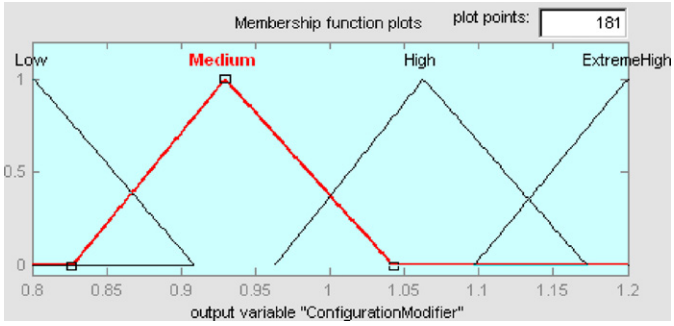

Fig. 9. Configuration membership function.

is changed significantly. Therefore, to define the configuration modifier, a range from 0.8 to 1.2 is selected, which results in a median of 1 and a maximum of a 50% increase (from 0.8 to 1.2). Four fuzzy sets are defined for the configuration modifier: low, medium, high, and extremely high, as shown in Fig. 9.

After the determination of membership functions, the fuzzy "if-then" rules were established. For any given input data for route-independent parameters, the corresponding modifier was derived from this fuzzy model. Fig. 10 illustrates the defuzzification process to derive the modifier for driver experience, as well.

4.4. The effects of number of lanes, truck configuration, population density, and road condition on the frequency

Figs. 11–13 and Table 5 present the effects of *number of lanes, truck configuration, population density,* and *road condition* on the frequency.

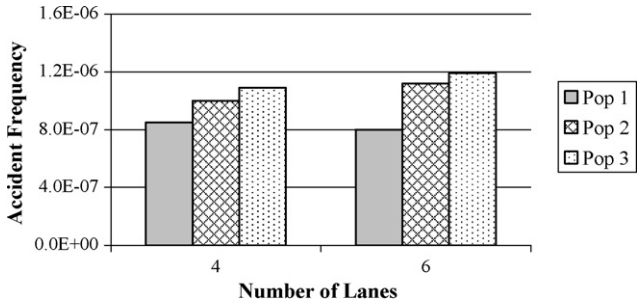

Fig. 11. Effects of number of lanes and population density on accident frequency.

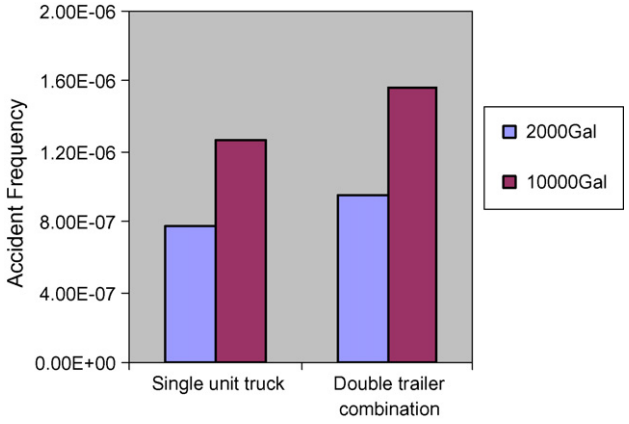

Fig. 12. Effects of container capacity and truck configuration on accident frequency.

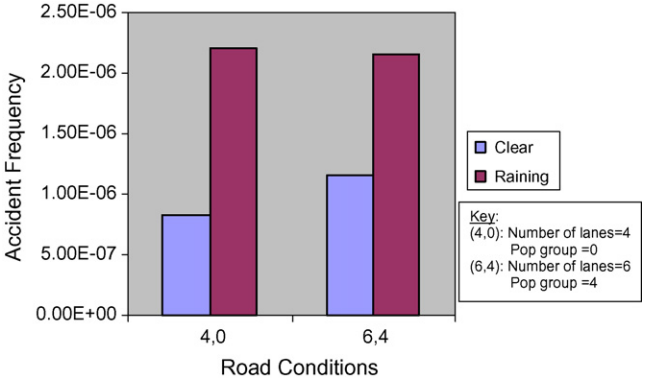

Fig. 13. Effects of weather conditions on accident frequency.

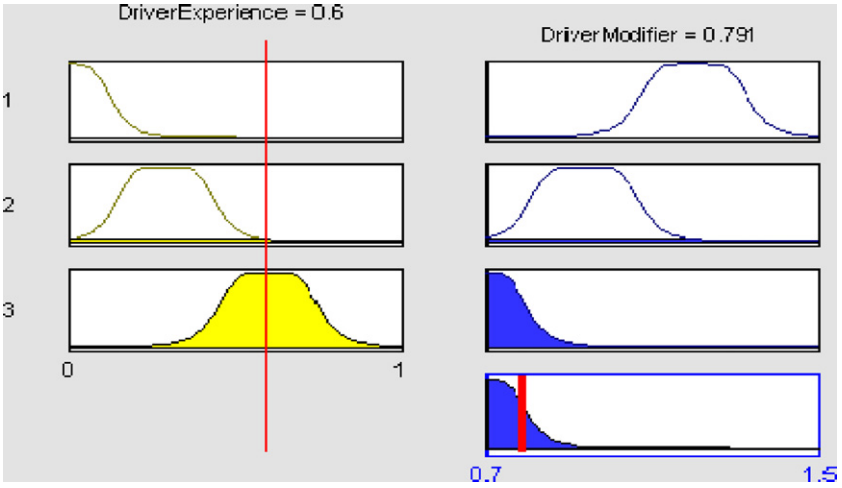

Fig. 10. Defuzzification process with Matlab.

Table 5
Effects of Driver experience on accident frequency.

Driver experience (normalized)	Driver modifier	Accident frequency (accident/vehicle mile)
0.1	1.14	1.65821E-06
0.6	0.791	1.15057E-06
0.8	0.86	1.25093E-06

Increases in population density and in number of lanes led to an increase in the frequency, as shown in Fig. 11. Also, as this figure indicates, in rural areas, the number of lanes did not impact accident frequency very much, i.e., when population density is enough lower, the number of lanes may not have a noticeable impact on accident frequency. Fig. 13 shows that weather conditions significantly affected frequency. Accident frequency increased under rainy weather in comparison to the frequency under clear weather. The increase was significantly higher when the number of lanes was six and the population group was between 5000 and 10,000.

As for route-independent parameters, an increase in both the complexity of vehicle configuration and container capacity resulted in an increased frequency (see Fig. 12). Fig. 4 illustrates the effects of driver experience. The accumulation of driving experience reduced the probability of an accident.

5. Summary

This paper presents a methodology, based on empirical data, to estimate accident frequency of HazMat transportation. The suggested integrated models incorporated the effects of both route-dependent and route-independent variables. Route-dependent variables included number of lanes, weather conditions, and population density. Route-independent variables were truck configuration, container capacity, and drivers' experience. Data from a variety of publicly available accident databases were utilized to establish a framework for estimating frequencies. Fuzzy logic was used to facilitate membership functions for the above six variables. Then, the methodology was implemented on a segment of highway US 290 in Texas, to establish frequency values for the six variables.

The proposed methodology provides fundamental information that is required to perform overall risk analysis along a route. The methodology was established as part of an effort to develop a HazMat transportation optimization procedure. It is important to note, though, that this model provides estimations only. However, the model can be used as a basis for the development of a methodology to predict potential accidents. These could include predictive models for the number, rate, frequency, and severity of accidents using past accident information, which then can be analyzed for sensitivity.

References

[1] Office of Hazardous Materials Safety, http://hazmat.dot.gov/hmpe_execsum.pdf, March 20, 2005.
[2] U.S. Department of Transportation, www.dot.gov, June 2, 2005.
[3] W. Rhyne, Hazardous Materials Transportation Risk Analysis: Quantitative Approaches for Truck and Train, Van Nostrand Reinhold, New York, 1994.
[4] Center for Chemical Process Safety, Guidelines for Chemical Transportation Risk Analysis, American Institute of Chemical Engineers, New York, 1994.
[5] D. Harwood, E. Russell, Present Practices of Highway Transportation of Hazardous Materials, FHWA/RD-89/013, Federal Highway Administration, DOT, Washington, DC, 1990.
[6] D. Brown, W. Dunn, A National Risk Assessment for Selected Hazardous Materials in Transportation, ANL/DIS-01-1, Argonne National Laboratory, Argonne, 2000.
[7] Battelle, Comparative Risks of Hazardous Materials and Non-Hazardous Materials Truck Shipment Accident/Incident, Prepared for Federal Motor Carrier Safety Administration, March 2001.
[8] R. Bubbico, S. Cave, B. Mazzarotta, Risk assessment for land transportation of hazardous materils in Italy, Loss prevention and safety promotion in the process industries 11th International Symposium Loss Prevention 2004 Praha Congress Centre, 4297–4304, 31 May–3 June, 2004.
[9] R. Bubbico, S. Cave, B. Mazzarotta, Risk analysis for road and rail transport of hazardous materials: a GIS approach, J. Loss Prev. Process Ind. 17 (2004) 483–488.
[10] J. Yen, R. Langari, Fuzzy Logic: Intelligence, Control, and Information, Prentice-Hall, Upper Saddle River, 1999.
[11] U.S. Census Bureau, http://www.census.gov/econ/www/cfsnew.html, June 2, 2005.
[12] Matlab Toolbox Manual.
[13] A. Cameron, P. Trivedi, Regression Analysis of Count Data, Cambridge University Press, 1998.
[14] National Climate Data Center, http://www.ncdc.noaa.gov/oa/climate/online/ccd/nrmpcp.txt, July 26, 2005.
[15] Athena Group, Rain Rates, http://my.athenet.net/~multiplx/cgi-bin/pics/rain_rate.html, July 26, 2005.
[16] M. Armstrong, A Handbook of Management Techniques: A Comprehensive Guide to Achieving Managerial Excellence and Improved Decision-Making. 3rd ed., London, UK, 2006.
[17] T.L. Saaty, Decision Making for Leaders, RWS Publication, Pittsburgh, PA, USA, 1999.