

# A comparative study of models for the incident duration prediction

Gaetano Valenti · Maria Lelli · Domenico Cucina

Received: 3 September 2008 / Accepted: 12 April 2010 / Published online: 7 May 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

## Abstract

**Purpose** This study is intended to investigate the reliability of different incident duration prediction models for real time application with a view to contribute to the development of a decision aid tool within the incident management process context where rough incident duration estimates are currently provided by traffic operators or police on the basis of their skill and past experience.

**Methods** Five predictive models, ranging from parametric models, to non-parametric and neural network models, have been considered and compared evaluating their capacity of predicting incident duration. The data set used in this study for developing and testing the prediction models includes 237 incident events and contains information about the incident characteristics, the personnel and equipment involved to clear the incident and the related response times, including the beginning and ending time of the incident.

**Results** Testing results have demonstrated that the proposed models are able to achieve good performance in terms of prediction accuracy especially for incidents with duration less than 90 min. This finding is partly due to the fact that the dataset has a relatively small number of severe incidents. Furthermore a linear combination of predictions from models was applied with negligible gain in accuracy.

**Conclusions** A deeper investigation is suggested for a future work to evaluate potential improvements from the application of other combination methods. Moreover each proposed model is able to reach best performance for incidents within a particular duration range. Thus a preliminary incident

classification scheme could be more convenient in order to select the more appropriate prediction model.

**Keywords** Incident duration · Duration prediction model · Statistical models · Regression models · Discriminant analysis

## 1 Introduction

When an incident occurs, the timely estimate of its duration assumes a key role in the overall incident management process. Specifically reliable incident duration predictions can help traffic managers in providing correct and essential information to road users, applying appropriate traffic control measures at or near the incident location and evaluating the effectiveness of the incident management strategies implemented.

In current practices, rough incident duration estimates are provided by traffic operators or police on the basis of experience and the known characteristics of the incidents such as the nature of the incident, the occurrence of injuries and fatalities, as well as the type and number of vehicles involved. The reliability of these practices is still unknown and largely depends upon the skill and experience of the operator.

Grounded on the existing scientific literature, this study intends to develop and compare the effectiveness of different prediction models suitable to estimate the incident duration in a real-time environment. The proposed prediction models incorporate variables that have the greatest influence on the incident duration and that can be practically obtained in real time as soon as the incident is detected and verified.

The incident data used in this study for developing and testing the prediction models have been supplied by the “Fiano” Trunk Management Centre of Autostrade per l’Italia Spa which is the biggest Italian motorway company.

These data, usually obtained from the incident scene and manually logged by the TMC operators in a database,

---

G. Valenti (✉) · M. Lelli  
ENEA (Ente per le Nuove Tecnologie, L’Energia e L’Ambiente),  
Rome, Italy  
e-mail: gaetano.valenti@casaccia.enea.it

D. Cucina  
Department of Statistics, La Sapienza University,  
Roma, Italy

contain information about the incident characteristics, the personnel and equipment involved to clear the incident and the related response times, including the beginning and ending time of the incident.

First, a statistical analysis of these incident data was conducted to investigate the factors that influence the incident duration with the scope to find out what variables are important for the prediction process. Both the ANOVA and Kruskal-Wallis analysis have been performed to measure and test the statistical significance of differences in incident duration for each of the explanatory variables.

Then, different predictive models, ranging from parametric (polynomial-type) models, to non-parametric and neural network models, have been considered and compared evaluating their capacity of predicting testing data.

This paper is organized as follow: a review of previous studies on incident duration prediction, aimed at obtaining insight in the strength and weakness of the many methods that have been developed up to now is presented in the next section. This is followed by the exploratory analysis of incident data collected by “Autostrade per l’Italia Spa” to identify critical variables associated with the incident duration. Next the construction and testing of five incident duration prediction models are reported, namely: Multiple Linear Regression (MLN), Prediction/Decision tree (DT), Artificial Neural Network (ANN), Support/Relevance Vector Machine (RVM) and K-Nearest-Neighbour (kNN). Finally some practical conclusions are drawn from the comparison of their prediction performance in the various incident situations.

## 2 Previous studies on incident duration prediction

Incident duration is the time elapsed from the incident occurrence until all evidence of the incident has been removed from the incident scene. Incident duration consists of three stages: Reporting, Response and Clearance time (Fig. 1). Reporting is the time between the incident occurrence and the determination of the precise location

and nature of the incident. Response is the time needed to dispatch the appropriate rescue personnel and equipment to the incident site. Finally Clearance is the period of time between the arrival of response units and the restoring of the roadway capacity to its pre-incident conditions.

Over the past two decades a number of studies have been undertaken to investigate the feasibility of estimating incident duration. Various approaches, ranging from statistical modeling methods, to machine learning methods like neural networks, have been applied. However, a direct comparison of the results of these studies is quite difficult since datasets, used to build and validate the various models, exhibit different characteristics, reflecting local variations in data collection and reporting practices.

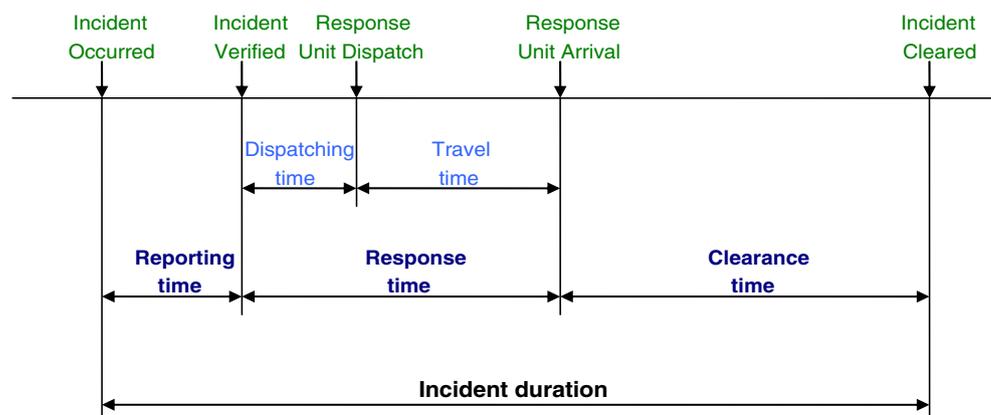
The purpose of developing incident duration models is to determine the relationships between incident duration and influencing variables. Previous studies reported similar sets of variables affecting incident duration, such as the incident type and severity, the number and type of vehicles involved, the geometric characteristics, the time of day and the emergency equipment (ambulances, tow truck, etc.) dispatched.

Golob, et al. (1987) [1] analyzed over 9,000 truck-involved accidents that occurred during a 2-year period on freeways in the greater Los Angeles area. Statistical models, that relate incident duration to collision type, accident severity and lane closures, were developed. The durations of incidents were found to be log-normally distributed for homogeneous groups of truck accidents, categorized according to the type of collision and, in some instances, the severity.

Also Giuliano (1989) [2] aggregated incidents into broad categories and estimated models as a function of incident characteristics for each category.

Jones et al. (1991) [3] introduced the important concept of conditional probability; that is, given that the incident has lasted  $X$  minutes, it will end in the  $Y$ th minute. The authors analyzed 2,156 incidents in the metropolitan Seattle area and found that the duration of incidents is approximated by a log-logistic instead of a log-normal distribution.

**Fig. 1** Components of incident duration



Ozbay and Kachroo (1999) [4] focused on incident having major impact on traffic and proposed the use of decision trees with the first split at the root node on the “incident type” variable. In this study a normal distribution of duration for homogeneous subsets of incidents (in terms of incident type and severity) was found.

Nam and Mannering (2000) [5] applied hazard-based duration models to statistically evaluate the time it takes to detect/report, respond to, and clear incidents. The model estimation results showed that a wide variety of factors significantly affects incident times, and that different distributional assumptions for the hazard function are appropriate for the different incident times being considered.

Smith and Smith (2001) [6] proposed and applied nonparametric regression and classification trees as models to predict incident clearance time.

Lin et al. (2004) [7] presented a system that integrates the discrete choice model with a rule-based supplemental module for estimating the duration of a detected incident. The primary function of the embedded discrete model is to estimate those incidents having durations less than 60 min. For severe incidents that may last more than 1 h, the system uses a rule-based supplemental module.

Wang et al. (2005) [8] developed two models to predict the vehicle breakdown duration: one based on fuzzy logic (FL) and the other on artificial neural networks (ANN). The study demonstrated that FL and ANN can provide reasonable estimates for the breakdown duration with few variables. However, both models had difficulties in predicting the outliers.

Ozbay and Noyan (2006) [9] used Bayesian Networks (BNs) as knowledge discovery process to accurately predict incident duration. The research showed that BNs offer an effective way to represent the stochastic nature of incident.

On the basis of these previous studies (see also [10, 11]), it can be concluded that each method seems to have its own strengths and weaknesses, thus no single method is expected to be the best method under all circumstances. If the full incident duration prediction horizon is to be covered, a combination of methods seems to be the best option. This view motivates the focus of this study on comparing different incident duration prediction methods.

### 3 Data description

The data used in this study are from the Incidents Database of “Autostrade per l’Italia Spa”, for two motorway sections, respectively of two and three lanes in both directions. They are referred to 3 months of 2005 (January, April and August) for the amount of 237 incident events.

These data are normally used for monitoring incident management operations and are related to every event

disrupting the regular traffic flow on the infrastructure by obstructing part of the road.

All the records of the database contain at least: 1) the starting and the ending time/date of the incident, 2) the type of the incident (crash, disabled vehicle, vehicle fire, obstacles on the road), 3) the location and the detection source.

The recorded information on the incidents can be divided into three different groups:

1. incident attributes (number of personal injuries/fatalities, number/type of vehicle involved, weather conditions, occurrence of events connected to the incident like cargo spill)
2. operational details (presence/number of emergency medical services, presence/number of special rescue vehicles...)
3. variables describing the state of the infrastructure and of the traffic (number/type of lane closed, queues...)

The statistical analysis showed that the incident durations distribution is right-skewed (skewness=1.73), as shown in Fig. 2: the mean value and the standard deviation are respectively of 45 min and 29 min. About 32% of the incidents has a duration of 30 min or less, whereas 78% of the incidents has a duration of less than 60 min. Only 8% of the incidents is longer than 90 min.

The Chi-Square statistical test confirmed that the incident duration distribution is a log normal distribution ( $p$ -value=0.053), as found by Giuliano [2].

Analysis of variance (ANOVA) was applied to determine which variable is statistically relevant for estimating incident duration. Moreover the non-parametric Kruskal-Wallis test was performed when the two assumptions of normality and homogeneity of variances, requested by the ANOVA, are not met.

Statistically significant differences were found only for 13 independent variables, listed in Table 1. Particularly the categorical variables with two possible values, such as

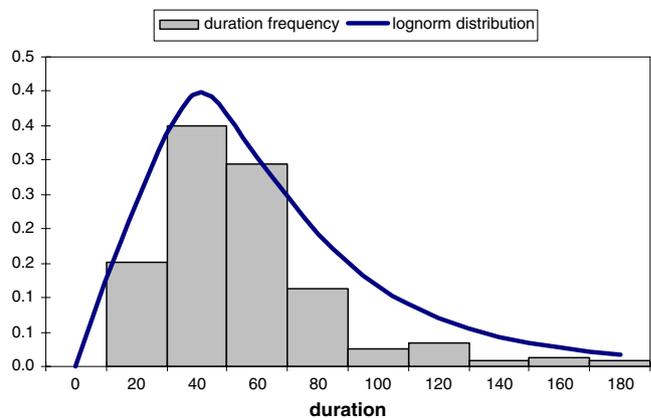


Fig. 2 Incident duration distribution

**Table 1** Possible independent variables resulted significant from ANOVA

Variable		Value
Incident type	Injuries/fatalities	1=Yes; 0=No
	Property damage (involving damage to the vehicle)	1=Yes; 0=No
	Disabled vehicle	1=Yes; 0=No
	Vehicle fire	1=Yes; 0=No
Incident details	Heavy duty vehicles involved	1=Yes; 0=No
	Peak hour	1=Yes; 0=No
	Infrastructure damage	1=Yes; 0=No
Operational details	Response service requested	1=Yes; 0=No
	Emergency Medical Services at the scene	1=Yes; 0=No
	Special agencies (heavy tow truck, HAZMAT clearance agency)	1=Yes; 0=No
Infrastructure and traffic variables	Number of lanes	1=3 lanes; 0=2 lanes
	Shoulder lane occupied	1=Yes; 0=No
	Slow lane occupied	1=Yes; 0=No

heavy duty vehicles involved, the presence of injuries and the presence of emergency service at the scene, resulted to be significant instead of the corresponding numerical variables (number of heavy duty vehicles involved, number of injuries, number of emergency services). Some variables that resulted significant for many models in literature, like weather conditions, were found not determinant in this study. The explanation of these apparently divergent results probably lies in the limited number of incident events with non-zero values for the related variables.

#### 4 Models to predict incident duration

There is a wide range of methods that may be applicable to incident duration prediction. In this study, five incident duration prediction models are discussed and compared, namely:

- 1) Multiple linear regression (MLR);
- 2) Prediction/Decision tree (DT);
- 3) Artificial Neural Network (ANN);
- 4) Support/Relevance Vector Machine (RVM);
- 5) K-Nearest-Neighbour (KNN).

For assessing the predictive ability of these models, the incident data set was split into training and testing partitions with statistical properties similar to those represented in the original dataset. Specifically, 187 incident cases were included in the training partition for the model construction process, whereas 50 incident cases were used to evaluate the accuracy of the proposed models.

Moreover, four incident duration classes were used to estimate and compare the models performance at the different duration horizons according to the incident severity: short (<30 min), medium (31–60 min), medium-long (61–90 min) and long (>90 min).

For investigating the accuracy of the proposed models the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE) and the Mean Absolute Percentage Error (MAPE) were adopted. The MAE quantifies the average magnitude of the errors, the RMSE diagnoses their variation and the MAPE weights them in relation to the actual value amount.

##### 4.1 Multiple linear regression

Multiple linear regression attempts to model the relationship between two or more independent or explanatory variables ( $X_1, X_2, \dots, X_p$ ) and a dependent variable ( $Y$ ) by fitting a linear equation to observed data ([12, 13]).

In this study linear regression with the  $\log_{10}$  of the incident duration as the dependent variable was used in order to meet the normal distribution assumption required by the MLR method. The skewness coefficient for the  $\log_{10}$  distribution was equal to  $-0.24$ .

Next the step-wise approach was adopted for determining the independent variables that should be included into the models. The resulting best-fitting model consisted of 6 independent variables plus a constant term— $\log_{10}(\text{duration}) = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6$ —with the independent variables and related coefficients, reported in the following Table 2:

All coefficients are statistically significant at the 95% level, however the explanatory power of the model is rather poor as indicated by  $R^2=0.32$ . Furthermore the F ratio is equal to 7.545 and the p-value is equal to zero. The addition of other variables does not significantly improve the accuracy of the predictions. The MAE value is 17 min.

Figure 3 shows the results of applying the MLR model to the testing data set. From this figure, it can be seen that the model tends to underestimate the durations for the incident cases with higher durations partly because the data set used has relatively small number of severe incidents; the five incident cases with the longest durations have an absolute error higher than 30 min. Moreover 33 out of 50 incident cases have an absolute error less than 20 min, while 45 out of 50 incident cases have an absolute error less than 30 min. The MLR model is more accurate in predicting short duration incident cases where the MAE value is 9 min.

The MAE value achieved by the MLR model is comparable to the MAE obtained by Ozbay and Kachroo [4] with DT models, and better than the one got by Smith & Smith [6] using KNN models.

**Table 2** Coefficients of the MLR model

	Variable	Coefficients (b <sub>i</sub> )	Std. Error	t-student	p-value
a	Constant term	1.657	.092	17.924	.000
X <sub>1</sub>	Emergency Medical Services at the scene	.222	.040	5.617	.000
X <sub>2</sub>	Heavy duty vehicles involved	.147	.039	3.820	.000
X <sub>3</sub>	Peak hour	.156	.035	4.488	.000
X <sub>4</sub>	Infrastructure damage	.204	.078	2.619	.010
X <sub>5</sub>	Number of lanes	-.092	.035	-2.657	.009
X <sub>6</sub>	Vehicle fire incident	-.276	.088	-3.148	.002

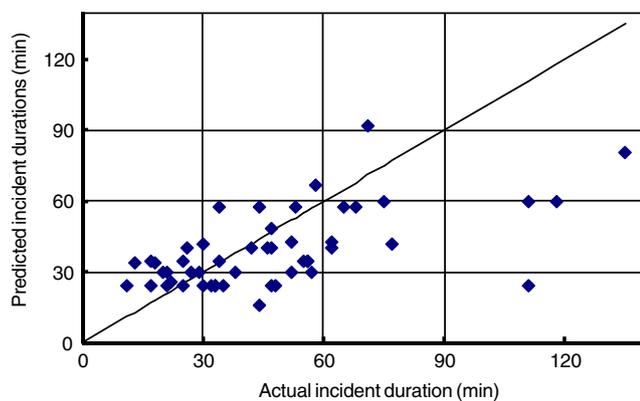
### 4.2 Prediction/Decision tree

Prediction/Decision trees can perform classification for predicting what group a case belongs to, as well as regression for predicting a specific value. DTs are non-parametric models as they make no assumption on the data distribution and, as a result, they may be applied in situations where little is known about the application in question.

As with all regression techniques we assume the existence of a single output (response) variable and one or more input (predictors) variables. It is called a decision tree because the resulting model is presented in the form of a tree structure or a set of logical “if-then” conditions (tree nodes). The visual presentation makes the decision tree model very easy to understand and assimilate.

Decision tree is built through an iterative process of splitting the data into partitions, and then splitting up further on each of the branches. The process continues until each node reaches a user-specified minimum node size and becomes a terminal node. The terminal nodes of the tree contain the predicted output variable values. The theoretical and computational details of decision tree model are provided in [14–17] and [18].

In this study, the statistical CHAID (chi-squared automatic interaction detection) procedure was used to iteratively



**Fig. 3** MLR prediction results versus duration test data

segment the incident data set into mutually exclusive subgroups according to the explanatory power of a set of predictors with regard to the incident duration. The application of CHAID procedure to the incident data set allowed to identify the four variables (Heavy duty vehicles involved, Emergency Medical Services at the scene, time of day, and number of lanes) that are most influential predictors of incident duration. The results of CHAID procedure are illustrated by the tree diagram reported in Fig. 4.

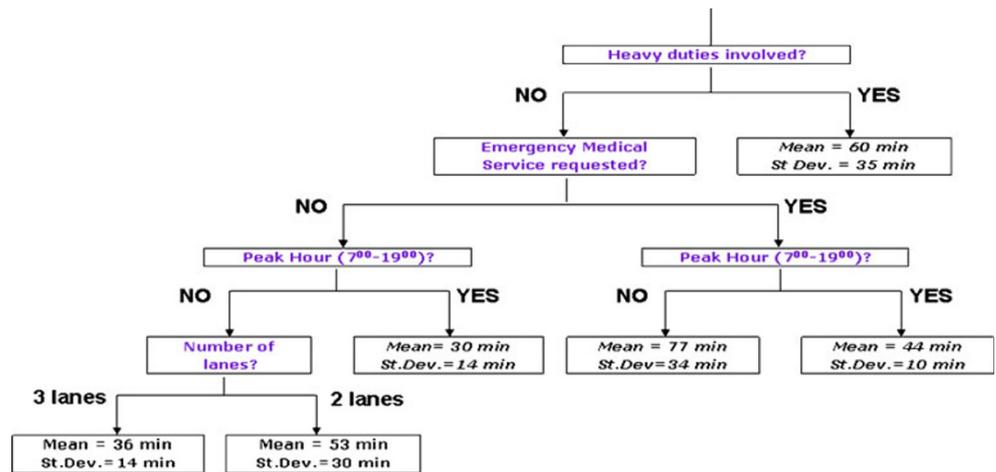
Validation test results showed that the developed DT model has satisfactory precision in predicting the duration of most incident cases. In particular 37 incident cases out of 50 are predicted with less than 20 min of prediction error. Better prediction performance is given by the DT model for incident cases with medium-long durations, where the MAPE and MAE values are equal to 18% and 12 min, respectively.

### 4.3 Artificial neural network

An Artificial Neural Network (ANN) model is a flexible mathematical structure capable of describing complex nonlinear relations between input and output datasets. ANNs have been successfully applied to prediction and pattern classification problems [19]. The architecture of ANN models is loosely based on the biological neural system. Although there are numerous types of ANNs, the most commonly used type of ANN is the Multi-Layer Perceptron (MLP). This is a feed-forward, fully-connected hierarchical network typically comprising three types of neuron layers each including one or several neurons: an input layer, one or more hidden layers and an output layer. The behaviour of a neural network is determined by the transfer functions of its neurons, by the learning rule and by the architecture itself ([20, 21]).

In this study, the number of neurons in the input layer is determined by the 13 most significant variables affecting incident duration, while a single neuron in the output layer is made up of the incident duration value being predicted. Moreover various ANN architectures, with one or two hidden layers and different number of neurons in the hidden

**Fig. 4** Decision tree for incident duration prediction



layers, were trained using the Levenberg-Marquardt back-propagation algorithm.

The best performing ANN architecture is obtained with a single hidden layer of 15 neurons and employing tangent-sigmoid transfer functions.

The MAE value for this ANN model is equal to 17 min, and 32 out of the 50 incident cases have been predicted with an absolute error less than 20 min. In Fig. 5 the ANN model results are reported versus the actual durations. The ANN model has a satisfactory accuracy for the incident cases with duration longer than 60 min. However the ANN model tends to overestimate the prediction values for the short duration incident cases (0–30 min), where the MAE value is equal to 18 min.

4.4 Support/Relevance vector machine

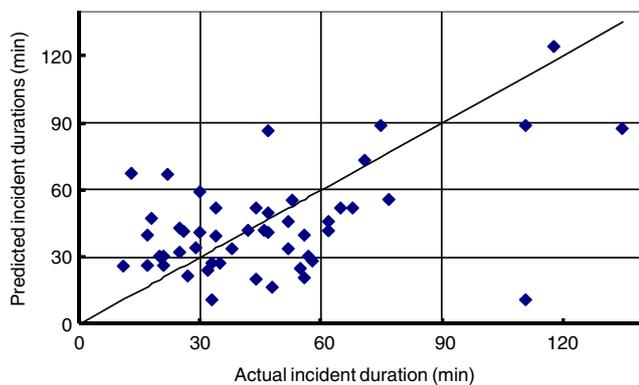
The Support Vector Machines (SVMs) are supervised learning machines born in the 1990s in the framework of statistical learning theory, based on the Structural Risk Minimization Theory (SRM) developed by Vapnik and Chervonenkis [22], to clarify the properties of generalization of the learning machines. The SVMs are powerful tools for solving problems of classification, regression, pattern

recognition, density estimation [23], with the supervisor’s output as a function of a linear combination of kernel functions centred on a subset of the training data, consisted of the so called support vectors.

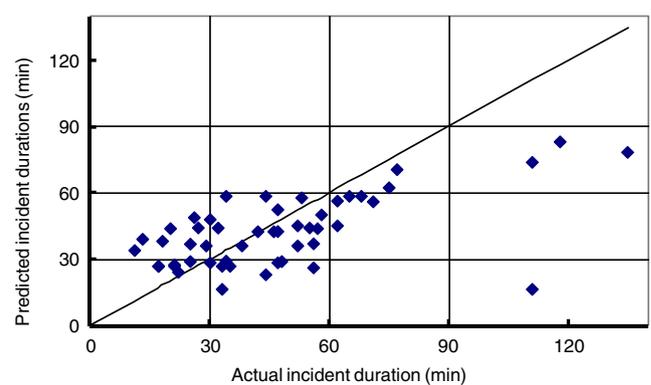
In the last years many different SVM models were developed, based on a variety of error functions, or kernels or optimization techniques. In 2001 Tipping [24] elaborated a new support vector machine, called Relevance Vector Machine (RVM), merging the Vapnik theory with the Bayesian statistics. The RVM model is based on a hierarchical prior on the parameters of the kernel functions’ weights, which leads to model sparseness. As a consequence, the RVM can generalize well and can provide inferences at low computational cost, bypassing some SVM constrains.

In this study, different SVMs and RVMs were trained varying kernel and error functions with different set of independent variables. Using the Cauchy Function Kernel and the training dataset composed of the 13 significant explanatory variables from ANOVA, the best performing RVM was obtained with 45 support vectors. This model gave the smallest MAE of about 15 min.

The validation test results, reported in Fig. 6, demonstrate that RVM model provides good performance in



**Fig. 5** ANN prediction results versus duration test data



**Fig. 6** RVM prediction results versus duration test data

predicting durations for incident cases belonging to the medium and medium-long duration groups. The MAEs for these two groups are equal to 12 and 10 min, respectively. In particular for the medium-long duration incident cases, the MAPE is rather small (15%). Only for the four longest duration incidents the prediction absolute errors are greater than 30 min, while 38 out of the 50 incident cases have an absolute error less than 20 min.

#### 4.5 K-nearest neighbour

The non-parametric K-Nearest Neighbour (KNN) method offers an alternative to the traditional parametric regression models. Through this method, the estimate/prediction for a current observation is simply based on weighting the contributions of the *k* nearest neighbours, so that the nearer neighbours contribute more than the farther ones.

The neighbourhood size is defined using independent variables which are known in both the past and current observations. In order to define the relative closeness of a given point, the form of the similarity (or distance) measure must be specified. Similarity measures based on absolute differences or Euclidean distance functions are typically applied.

In building the KNN model the choice of *k* can strongly influence the quality of predictions: a small value of *k* leads to a large variance in predictions; alternatively, setting *k* to a large value may lead to a large model bias since the *k* nearest neighbours are farther away including cases that are less representative of the case under examination. Thus, *k* should be set to a large value enough to minimize the estimation error and small enough (with respect to the numbers of cases) so that the *k* nearest points are close enough to the query point.

In this study an appropriate distance metric, based on the number of matching independent variables between past and current incident, was applied since all the independent variables are binary (0/1) in form [6]. Furthermore weight factors for each independent variable, given by the absolute difference between the average duration of the two related yes/no samples, were used to compute the KNN distance. *K* values up to 30 were tested and compared, using the MAE as measure of effectiveness. The minimum value of MAE was obtained in correspondence of *K*=10.

The KNN model results for the 50 accidents in the testing set are illustrated in Fig. 7. The error between predicted and actual incident durations averages over 17 min. Slightly more than half of the predicted durations are within 15 min of the actual time. The KNN model usually tends to overestimate the prediction values for incidents of relatively short duration, while underestimating them for incident cases of relatively long duration.

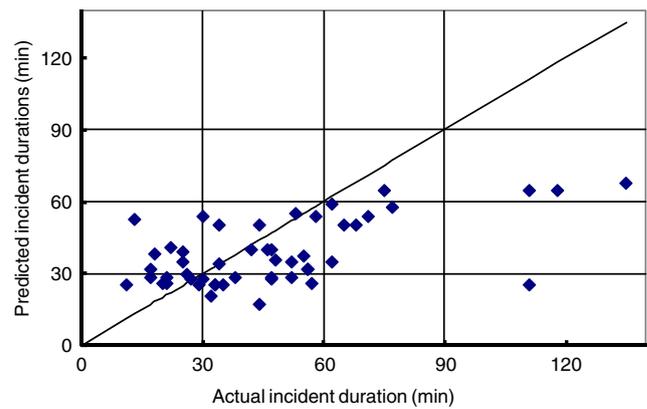


Fig. 7 KNN prediction results versus duration test data

### 5 Comparisons and conclusions

This paper presents the findings of a study that appraises and compares five predictive modelling methods, ranging from parametric (polynomial-type), to non-parametric and neural network models in order to provide an useful and reliable decision aid tool within the incident management process context where rough incident duration estimates are currently provided by traffic operators or police on the basis of their skill and past experience.

These models have been developed and tested using a common incident data set, including 237 incident events, for allowing a direct comparison of the models' prediction ability in the various incident situations. The Mean Absolute Errors (MAE), the Root Mean Squared Errors (RMSE) and the Mean Absolute Percentage Errors (MAPE) were adopted to estimate the models' accuracy.

The testing results, based on 50 incident events, have demonstrated that the proposed models are able to achieve good performance in terms of prediction accuracy for incidents with duration less than 90 min, matching what was obtained in past studies using similar prediction methods.

As reported in previous studies, the models' prediction ability is heavily affected by the quality of the input data. All the applied methods get the worst performances (Table 3) in two incident cases having durations incoherent with the variables' values that describe the characteristics of

Table 3 Worst prediction errors (prediction-incident duration)

Incident type	Incident duration (min)	MLR error (min)	CHAID error (min)	ANN error (min)	RVM error (min)	KNN error (min)
HDV involved	13	20,73	46,67	54,36	25,99	39,48
Obstacles on the road	111	-86,96	-81,33	-100,38	-94,61	-85,88

the incidents. This incoherence is most likely due to errors occurred in the logging of the incident data.

By excluding the two incident cases reported above the model indicators were further estimated (Tables 4 and 5).

According to the MAE and RMSE values in Table 4, the less reliable model is the decision tree model (CHAID), characterised by the greatest variance in the errors. The RVM is the most reliable model performing the smallest MAE and RMSE values, while the smallest MAPE value is performed by the MLR, working with only six explanatory variables. The highest MAPE is achieved by the ANN, with great errors for short duration cases, as shown in the previous section.

As listed in Table 5, 79% of RVM prediction errors are less than 20 min, while CHAID predictions are with the greatest number of errors more than 30 min. However the CHAID model exhibits the greatest number of errors less than 5 min, and this result is achieved with only four explanatory variables. The advantage to give a ready-to-use easy tool, with a small number of variables, makes the Decision Tree together with the MLR the methods most used for the incident duration prediction problem. Moreover, unlike “black box” methods such as ANN and RVM, a further advantage of these methods is their statistical approach that allows a transparent and easy-to-understand explanation of their results.

A further step to enhance the prediction accuracy using all the proposed models can be a combination procedure of their predictions, in order to exploit the fact that the models have strengths and weaknesses in different situations. In this view, the linear combinations proposed by Granger and Ramnathan [25] were applied. From this application a negligible gain in prediction accuracy was reached in terms of MAE values (from 13,65 with RVM to 12,62 min with predictions’ combination). However a deeper investigation can be suggested for a future work to evaluate potential improvements from the application of other combination methods.

Moreover, looking at the MAEs calculated for the four duration classes (Fig. 8), the following findings can be highlighted:

- the MLR is the best performing model for short duration incidents;

**Table 4** Number of variables, MAE, RMSE and MAPE of the models

	MLR	CHAID	ANN	RVM	KNN
N° variables	6	4	13	13	10
MAE	15,17	<b>16,66</b>	16,02	<b>13,65</b>	15,41
RMSE	20,04	<b>23,07</b>	19,80	<b>17,29</b>	20,29
MAPE	<b>34%</b>	43%	44%	36%	36%

**Table 5** Distribution of predictions’ absolute errors for all models

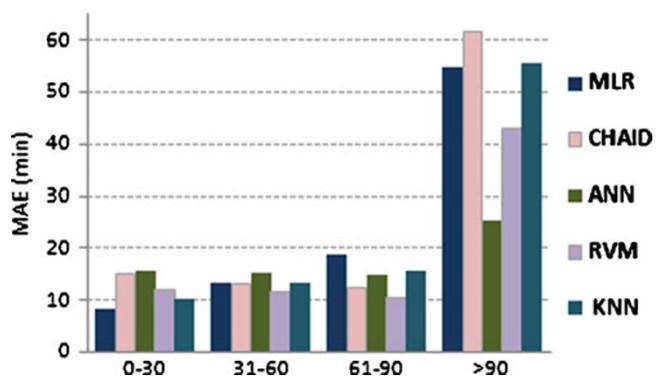
Absolute error (min)	MLR	CHAID	ANN	RVM	KNN
<5	19%	<b>27%</b>	13%	17%	19%
5–10	29%	19%	29%	31%	23%
10–20	21%	33%	25%	31%	35%
20–30	23%	8%	21%	15%	15%
>30	8%	<b>15%</b>	13%	6%	8%

- the best predictions are achieved by the RVM model in the incident cases with medium/medium-long duration;
- the ANN is the only model that can predict an incident longer than 90 min. Moreover the ANN model gives the best results for long duration incident cases, with the lowest MAPE, even if greater than 30%;
- all proposed models tend to have a relatively low accuracy for incidents with long duration partly because the dataset has a relatively small number of severe incidents.

In conclusion, each proposed model is able to reach the best performance for incidents within a particular duration range, as if they have specialised skills in predicting incidents of specific duration class. For this reason, a preliminary incident classification scheme would be more convenient in order to select the more appropriate prediction model. For example, a preliminary classification between two classes of duration—less or more than 30 min—can help to pick up between the MLR for short duration incidents and RVM for the others.

Finally the findings reached in this study have certainly demonstrated the validity of the RVM as prediction model also in the context of incident duration prediction.

However it is likely the proposed models could have a limited accuracy when used in other geographical contexts where different incident management and emergency response actions take place. In order to ensure that the proposed models are able to deal with different



**Fig. 8** MAE values achieved by the proposed model

conditions, a wider-scale data collection effort is needed to be undertaken.

**Acknowledgment** The authors would like to thank Autostrade per l'Italia Spa for kindly supplying the incident data used in this study.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Golob TF, Recker Leonard JD (1986) An analysis of the severity and incident duration of truck-involved freeway accidents, Institute of Transportation Studies, Univ. of California, Irvine, UCI-ITS-WP-86-9
- Giuliano G (1989) Incident characteristics, frequency, and duration on a high volume urban freeway. *Transp Res* 23A (5):387–396
- Jones B, Janssen L, Mannering F (1991) Analysis of the frequency and duration of freeway accidents in seattle. *Accident Anal Prev* 23(4):239–255
- Obzay K, Kachroo P (1999) Incident management in intelligent transportation systems. Artech House, Boston
- Nam D, Mannering F (2000) An exploratory hazard-based analysis of highway incident duration. *Transp Res, Part A Policy Pract* 34(2):85–102
- Smith BL, Smith K (2001) Forecasting the clearance time of freeway accidents. Research Report No. UVACTS-15-0-35. USDOT University Transportation Center
- Lin PW, Zou N, Chang GL (2004) Integration of a discrete choice model and a rule-based system for estimation of incident duration: a case study in Maryland. The 83rd Annual meeting of the Transportation Research Board
- Wang W, Chen H, Bell MC (2005) Vehicle breakdown duration modelling. *J Transp Stat* 8(1):75–84
- Ozbay K, Noyan N (2006) Estimation of incident clearance times using Bayesian Networks approach. *Accident Anal Prev* 38 (3):542–555
- Wei C-H, Lee Y (2007) Sequential forecast of incident duration using Artificial Neural Network models. *Accident Anal Prev* 39 (5):944–954
- Garib A, Radwan AE, Al-Deek H (1997) Estimating magnitude and duration of incident delays. *J Transp Eng* 123(6):459–466
- Johnson R, Wichern D (2003) Applied multivariate statistical analysis. Prentice Hall
- Fox J (1997) Applied regression analysis, linear models, and related methods. Sage
- Brieman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth International Group, Belmont
- Kass GV (1980) An exploratory technique for investigating large quantities of categorical data. *Appl Stat* 29:119–131
- Loh WY, Shih YS (1997) Split selection methods for classification trees. *Stat Sinica* 7:825–840
- Biggs D, DeVille B, Suen E (1991) A method of choosing multiway partitions for classification and decision trees. *J Appl Stat* 18(1):49–62
- Quinlan JR (1993) C4.5: Programs for machine learning. Morgan Kaufmann, San Mateo
- Duda RO, Hart PE, Stork DG (2000) Pattern classification (2nd ed). John Wiley & Sons
- Haykin S (1999) Neural Networks: a comprehensive foundation. Prentice Hall
- Bishop C (1995) Neural networks for pattern recognition. Clarendon, Oxford
- Vapnik V (1999) An overview of statistical learning theory. *IEEE Trans Neural Netw* 10(5):988–999
- Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines. Cambridge University Press, and reference therein
- Tipping M (2001) Sparse Bayesian learning and the relevance vector machine. *J Mach Learn Res* 1, pp. 211–244 (<http://www.miketipping.com/index.php?page=rvm>)
- Granger CWJ, Ramanathan R (1984) Improved methods of combining forecasts. *J Forecast* 3:197–204