Creating predictive models for forecasting the accident rate in mountain roads using VANETs

Borja Bordel[1,2], Ramón Alcarria[1,2], Gianluca Rizzo[1], Antonio Jara[1]

[1] Institute of Information Systems. University of Applied Sciences Western Switzerland (HES-SO) Sierre, Vallais, (Switzerland)
[2] On leave from: Universidad Politécnica de Madrid, Madrid, España
bbordel@dit.upm.es; ramon.alcarria@upm.es; gianluca.rizzo@hevs.ch; jara@ieee.org

Abstract. Monitoring the road network status of an entire country in a visual way (as traditionally) is very hard, so different mechanisms to do it in an automatic manner have been investigated. In particular, nomadic pervasive sensing platforms based on VANETs have been recently deployed. However, the level of road damage is a relative variable, and it is necessary to predict the particular impact of the same in each case, in order to prioritize the conditioning works. Therefore, in this paper a predictive model for forecasting the accident rate in mountain roads, considering the measures previously obtained through a nomadic sensing environment (and through the weather office) is defined. The model considers the type of road under study as well as different analysis scales to perform the calculations. The model is based on Taylor's series and multivariate functions. Real data related to Valais (Switzerland) road network is employed to construct and validate the proposed model.

Keywords: Vehicular Ad-Hoc Networks; prediction models; data analysis; pervasive sensing

1 Introduction

Switzerland is the most mountainous country in Europe. Thus, many settlements, monuments and infrastructures are only accessible by road. In fact, roads account for roughly 90% of total transportation space in Switzerland [1]. This percentage, besides, is continuously growing, as during the last thirty years roads have gained around 15% of extra space. In particular, in the canton of Valais are located the highest mountains of all Switzerland, being this region separated from its neighbors by mountain ranges crossable by high mountain roads that are closed almost the entire winter [2]. However, in summer, weather is warmer and temperatures reach the thirty-five degrees in some occasions.

As a consequence, in Valais, motorways only occupy 9% (approximately) of space reserved for roads [1]. Besides, both motorways and other roads (see Fig. 1) are subject to very variable climatic conditions, which causes rapid (and potentially serious) damage to these infrastructures.

The orography and the high density of roads make very difficult to monitor the status of the Valais road network in a traditional way (visually, by means of workers

on the roads assessing their status). Thus, during the last years, different technologies related to Smart City solutions have been implemented in order to address this problem. Specifically, in Valais, the Institute of Information Systems (University of Applied Sciences Western Switzerland) together with the Mobility Lab in Sion have developed a Nomadic Scalable Ecosystem (NOSE) for pervasive sensing, computing and communication [3], which monitors the road network status through sensor nodes embedded into public buses (managed by La Poste Suisse). Those buses define a Vehicular Ad Hoc Network (VANET) which offers the obtained results to the stakeholders [4].

Fig. 1. Considered types of roads **(a)** Motorway **(b)** Urban road **(c)** Mountain road

Data obtained from this environment, nevertheless, cannot be directly employed by road management offices, as road damage is a relative variable: the most urgent conditioning works are those which prevent the highest number of accidents. It would be desirable, then, to predict the future accident rate (or dangerousness index, more properly) of a road, considering the obtained data through NOSE.

Therefore, the objective of this paper is to create a predictive model for forecasting the dangerousness index of Valais roads, considering the obtained data through NOSE. Besides, weather information from the Meteosuisse office was included in order to improve the prediction accuracy. In order to do that, the different types of roads were studied, and different analysis scales were considered (roads were divided into sections of 1 km, 2 km, etc.). Once correlations among the available data were located, Taylor's series and multivariate functions were employed to define the model.

The rest of the paper is organized as follows: Section 2 describes the state of the art on predictive models. Section 3 describes the main proposal, and Section 4 validates the performance of the model. Finally, Section 5 presents some conclusions.

2 State of the art on predictive models

Predictive models are a key technique in several fields. From medicine [5] to Cyber-Physical Systems [6], predictive models allow improving the systems' performance.

In relation to traffic research, most works are focused on noise prediction [7]. Different mathematical expressions (usually logarithmic functions), depending on the environment under study and the evaluated infrastructure are proposed, in order to preview the acoustic impact of roads. Other important group of works deals with the

idea of forecasting the traffic congestion. In particular, different mobility models are proposed depending on the urban or road configuration [8]. Besides, seasonal behaviors are also modeled in order to preview bottlenecks in the road network [9]. Other articles define statistical distributions and probabilistic models in order to predict collision between animals and vehicles [10]; or the "en route" mental overload [11].

Recently, besides, modern data analytics [12], stochastic modeling [13] and adaptive control [14] techniques have been also employed to define predictive model for traffic flow, especially in the context of smart cities and cyber-physical systems [15]. Predictive models for forecasting the traffic evolution at real-time based on Big Data technologies [16] have been reported as well; and application-specific predictive models to be used in key scenarios (such as critical situations [17]) may be also found. As relevant contributions, different simulators [18] and case studies about the use of traffic predictive models (e.g. Montreal [19], Taiwan [20], etc.) have been published too.

Finally, predictive traffic models usually are hierarchical and consist of various layers, where lower levels manage correlation among physical parameters and more abstract levels are employed to predict human and long-term behaviors [21]. Although modern learning techniques are commonly employed [22] in predictive models, pure mathematical models allow a better understanding of the relations among the different variables. In particular, in this work, Taylor's series and multivariable functions are employed.

3 Proposal

In this section, the creation of the predictive models is discussed. The creation process includes three steps: (i) data acquisition and pre-processing, (ii) correlation analysis; and (iii) model definition and particularization. The following subsections describe in detail each one of these steps.

3.1 Data acquisition and pre-processing

The deployed NOSE infrastructure monitors around 100 kilometers of roads in the canton of Valais, in the districts of Sion and Hérens. Three types of roads are monitored: the A9 motorway between Saxon and Sion (30 kilometers); urban roads in Sion (20 kilometers) and a mountain road between Sion and Arolla (40 kilometers). Fig. 2 shows this itinerary. This infrastructure offers data about location, time, vehicle (bus) speed, road temperature, water-film height, percentage of ice, friction, dew point and general road conditions (dry, wet, moisture, melted snow, snowy/icy).

In order to obtain a deeper knowledge about the situation, information from the Meteosuisse office maybe also considered. Meteosuisse offers measures about 145 different variables (approximately) with time precision of milliseconds. For this work, however, we are considering only four of these variables: the total thickness of recumbent snow (in centimeters), rainfall (daily sum, in millimeters), duration of

sunshine (hourly sum, in minutes) and the hourly solar radiation (in watts per square meter).

Data about the traffic statistics (including accidents) in Valais were also obtained. This information was employed to calculate the dangerousness index (*DI*) of the roads under study during the second half 2016. Basically, the dangerousness index of a road (1) is a relative measure of the number of accidents in this road (N_{acc}), considering the corresponding level of risk exposure (R, i.e. the traffic volume in vehicles·km). In order to remove random variations, temporal and spatial mean data are usually employed.

$$DI = \frac{N_{acc}}{R} \left(\frac{accidents}{vehicles \cdot km} \right) \tag{1}$$

A spatial database is used to store these parameters. A PostgreSQL 9.6 database was used, with PostGIS 2.3.2 extension for enabling spatial operations. Data provided by the sensor infrastructure is stored considering the timestamp for the data and the location where the data was generated, storing the values in *timestamp with timezone* and *geometry* data types. This enables the possibility of performing temporal requests (data ranges depending on dates) and spatial requests (data filtering by proximity or density).

Data coming from Meteosuisse is stored in two interrelated tables. The first one stores metadata information from weather stations, such as location, metrics and time of the last update. The second one contains observations related to the *station_id*. Each observation had a unique identifier consisting in an auto-increment number, acting as table key.

Data about traffic statistics were also imported to the spatial database. The table describing accidents contained accident coordinates (expressed in *geometry*) and ID of the road, time of the accident (*timestamp with timezone*), type, description, caused property damage, and type of injuries (death, major, and minor).

Fig. 2. Itinerate under study showed on the GIS tool

For information representation in a spatial way we used QGIS framework version 2.18.6. We obtained information in *vector* format from the PostGIS database and represented it as three layers. We used Microsoft Bing as base map for the overlays. The itinerate under study showed on the GIS tool can be seen in Figure 2.

We have presented the information graphically, and using maps of the region to make a selection of data according to the type of road that we defined in the introduction.

For this, we have obtained road sections of 500 m, 1 km and 2 km and we have defined for each section the type of road (motorway, urban road, and mountain road). For the measurement of the section we have used the measurement tools provided by QGIS. For each section, a data set has been generated with information provided by the sensor infrastructure, information generated by Meteosuisse from the nearest weather station (*meteosuisse.station.sion*), and the number and type of accidents that occurred on those road sections. This is the input information that we introduce in our correlation analysis process, described in Section 3.2.

3.2 Correlation analysis

Performing a complete correlation analysis would imply to define a clustering method based on a sliding window with a fix length (2 km, 1km or 500 meters in our proposal). This window should be moved through the entire itinerary under study, which should be also divided into cells (inside which it is suppose road conditions do not change). Cell by cell, the sliding window advances generating a set of set of measures whose geographical mean has to be calculated in order to remove random effects. Mean values, thus, would be employed to calculate correlations. This procedure, however, requires detailed information about traffic volume in each road section to be available.

Table 1. Traffic flow depending on the road typology

	Road type		
Vehicle · km (in hundreds)	Motorway	Urban road	Mountain road
	419	119	20

In this case, we are only provided with global information [2] (see Table 1), so it is proposed a simpler analysis method. Roads are divided into fixed consecutive sections of the specified length, which are employed to obtain the geographical means of parameters. With this information, related to the first half of the year, we are looking for correlation with the accident rate in the second half of the year (see Table 2 and Table 3). For this first work, besides, data acquisition time is not considered.

Variables whose correlation with the accident rate is under $|r| = 0.3$ are not considered relevant for this study. In that way, as can be seen, different variables may be relevant depending on the considered analysis scale and type of road (i.e. there is a strong dependency on the road configuration). In general, however, 1 km road sections present a good balance and pretty acceptable results. In relation to climatic

information, as only one measuring station is placed around the study area, temporal (monthly) correlations are obtained.

Table 2. Correlation between the accident rate and NOSE parameters

Road type	NOSE Parameters							
	Altitude	Speed	Temp	Waterfilm	Ice	Friction	Dew point	Road Cond
Motorway	-0.0294	0.082	-0.198	0.1295	---	-0.1118	-0.177	0.1504
	-0.0758	-0.434	0.086	-0.3894	---	0.5502	0.128	-0.371
	-0.1479	-0.325	-0.008	0.2677	---	0.2676	-0.081	---
Urban road	-0.3771	0.053	-0.3974	0.0435	---	-0.0186	-0.584	0.0291
	-0.352	-0.684	0.497	-0.5344	0.435	0.5444	-0.084	-0.556
	0.2640	0.232	0.3677	-0.0505	---	0.0256	0.6158	0.0186
Mountain road	-0.8648	0.099	0.777	-0.551	0.769	0.4929	0.813	-0.492
	-0.521	0.494	0.248	-0.593	0.531	0.491	0.431	-0.208
	0.6023	0.043	-0.580	-0.205	0	0.1168	-0.187	-0.381

Black numbers represent road sections of 2km, red numbers 1km sections and purple results 500m sections

Table 3. Correlation between the accident rate and Meteosuisse parameters

Road type	Meteosuisse Parameters			
	Snow	Rainfall	Sunshine	Radiation
Motorway	-0.4205	-0.0155	-0.1402	-0.1991
Urban road	0.1526	-0.2699	-0.4000	-0.4459
Mountain road	-0.4410	-0.4692	0.5481	0.5076

3.3 Model definition

As can be seen, correlations between the relevant variables identified in the previous subsection and the accident rate are, in general, around $|r| = 0.5$. This means that none of these variables is a direct cause of accidents, but they are involved in the accident occurrence. A combination of several of these variables, then, should explain the accident rate during the next temporal period (6 months). However, how these variables are weighted and/or combined in order to cause or not an accident is unknown. In this work, we are supposing the multivariate function $F(\vec{x})$ represent this weighted combination. This function, moreover and as any other function, may be expressed using Taylor's series (2) centered in the origin.

$$DI = F(\vec{x}) = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{k_1 + \cdots + k_d = k} \binom{k}{k_1 \cdot \ldots \cdot k_d} \frac{\partial^k F(\vec{0})}{\partial x_1^{k_1} \cdot \ldots \cdot \partial x_d^{k_d}} \left(x_1^{k_1} \cdot \ldots \cdot x_d^{k_d} \right) \qquad (2)$$

Where $\vec{x}$ is a d-dimensional vector containing the relevant variables to be considered in the predictive model.

The model, besides, may be simplified if all constant values are aggregated (3). Furthermore, as computational solution cannot deal with infinite sums, the proposed Taylor's series should be approximated through the sum of the first q terms.

$$DI = \sum_{k=0}^{\infty} \sum_{k_1+\cdots+k_d=k} \lambda_{k_1,\ldots,k_d} \left(x_1^{k_1} \cdot \ldots \cdot x_d^{k_d} \right) \approx \sum_{k=0}^{q} \sum_{k_1+\cdots+k_d=k} \lambda_{k_1,\ldots,k_d} \left(x_1^{k_1} \cdot \ldots \cdot x_d^{k_d} \right) \qquad \textbf{(3)}$$

In that way, the proposed predictive model would be totally defined if the $\lambda_{k_1,\ldots,k_d}$ unknown parameters were obtained. These parameters are time independent, although an identical mathematical expression could be obtained if there were considered time dependent. This set of parameters may be calculated considering the data collections employed in the previous subsection, and solving the resulting equation system using the optimization of the mean square error technique.

4 Experimental validation

An experimental validation was designed in order to validate the proposal and analyze the performance of the proposed models. This section describes the creation of a particular predictive model using the previously presented information, and evaluates the validity of the obtained predictions trying to forecast the accident rate in the second half of 2016 (using information about the first half).

4.1 Model creation

The objective of this subsection is to define a predictive model to forecast the accident rate in mountain roads, considering road segments with a length of 1km. In order to create this first predictive model three variables from NOSE project were considered: the ice percentage, the waterfilm height and friction. Besides, one climatic variable was also taken into account: rainfall. In that way, four input variables were considered.

On the other hand, in order to create the model it is necessary to select the order of the mathematical expressions. For this first example, we selected a quadratic model ($q = 2$).

With these choices, the created model presents fifteen unknown parameters. In order to determine their value, thirty different segments of mountain roads were analyzed. Data about the accident rate in these segments and information about the NOSE project (as well as climatic information from Meteosuiss) were used to calculate these parameters $\lambda_{k_1,\ldots,k_d}$ (see Table 4).

Table 4. Model parameters

$\lambda_{0,0,0,0}$	$\lambda_{1,0,0,0}$	$\lambda_{0,1,0,0}$	$\lambda_{0,0,1,0}$	$\lambda_{0,0,0,1}$	$\lambda_{1,1,0,0}$	$\lambda_{1,0,1,0}$	$\lambda_{1,0,0,1}$
$1{,}27 \cdot 10^7$	$-3{,}1 \cdot 10^7$	$-1{,}8 \cdot 10^5$	$-4{,}8 \cdot 10^3$	$-4{,}6 \cdot 10^5$	$1{,}91 \cdot 10^7$	$118{,}46$	$0{,}4437$
$\lambda_{0,1,1,0}$	$\lambda_{0,1,0,1}$	$\lambda_{0,0,1,1}$	$\lambda_{2,0,0,0}$	$\lambda_{0,2,0,0}$	$\lambda_{0,0,2,0}$	$\lambda_{0,0,0,2}$	
$7{,}06 \cdot 10^4$	$2{,}33 \cdot 10^5$	$5{,}57 \cdot 10^5$	$5{,}82 \cdot 10^3$	$33{,}02$	$1{,}69 \cdot 10^4$	$89{,}1053$	

4.2 Predictions and results

A set of new road segments (control group) was employed to evaluate the performance of the proposed model. Using information from NOSE project and Meteosuisse the predicted accident rate for the second half of 2016 was obtained. Results were compared with the real information from the Traffic Office.

In order to evaluate the quality of the performed prediction the mean square error was calculated and evaluated. Table 5 presents the error associated with each road segment.

Table 5. Mean square error of predictions

Accidents per road segment	Less than one	1 – 2	Two or more	Total
Error	44.26%	17.50%	58.55%	20,19%

As can be seen, extreme situations are badly predicted. This fact is due to the employed data during the creation of the predictive model. As only information about six months was available, no representative number of samples of these situations was considered. Future works will consider more precise predictions as a higher amount of information should be available.

Any case, as the proposed model produces decimal numbers, while the number of accidents is an integer, it is necessary to perform a cast between both values. Three possibilities were considered: rounding to the nearest integer, truncation to the lower integer, and truncation to the upper integer.

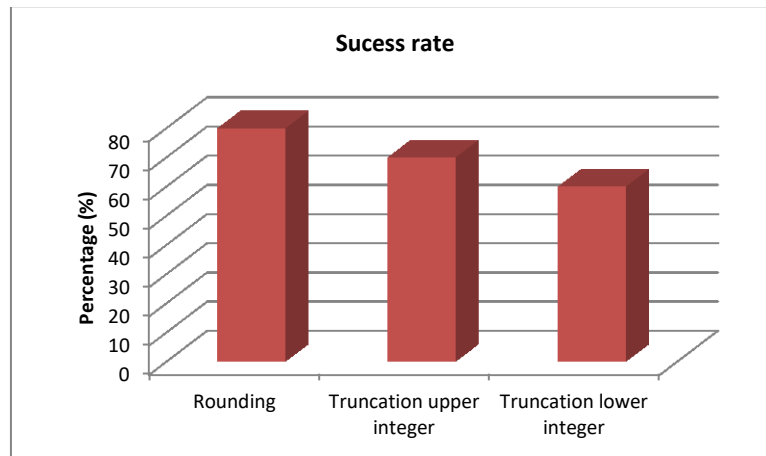Fig. 3 presents the resulting success rate in the predictions for each casting technique.

Fig. 3. Success rate depending on the casting technique

As can be seen, rounding technique is the most success solution. Around 80% of cases are predicted correctly. Truncation, however, presents more problems. In general, truncation to the upper integer is most precise (around 10% more) than truncation to the lower integer (which presents a success rate of 60%, approximately).

5 Conclusions

Traffic accidents are one of the most important social problems nowadays. Preventing or reducing the accident rate is a remarkable objective of most traffic management offices. In order to do that, roads have to be maintained in the best possible condition, however, resources are limited and decisions have to be made. In this paper it is proposed a predictive framework and model, in order to forecast the future accident rate in a road segment depending on a collection of variables monitored though VANETs. The proposed model is focused on mountain roads in the canton of Valais (Switzerland).

Results showed that the proposed model allows obtaining a general picture about the future accident rates, although detailed and more precise predictions require advanced solutions or more information about roads (currently, predictions present an error of 20%).

Future works should consider information acquired during longer time periods in order to remove all random phenomena, and more detailed geographical information about roads.

References

1. Federal Statistical Office (FSO). Land use in Switzerland. Results of the Swiss land use statistics Available online: https://www.bfs.admin.ch/bfsstatic/dam/assets/348992/master (accessed 26 august 2017)
2. Roads and traffic 2017 - Developments, facts, figures. Annual publication of the Federal Roads Office. Available online: https://www.astra.admin.ch/astra/en/home/documentation/facts-and-figures.html (accessed 26 august 2017)
3. Dupont, A., Bocchi, Y., & Rizzo, G. (2017, July). NOSE: A NOmadic Scalable Ecosystem for Pervasive Sensing, Computing and Communication. In International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (pp. 884-893). Springer, Cham.
4. NOSE dashboard. Available online: https://nose-e3fe0.firebaseapp.com/app/#/dashboard/sTemp (accessed 26 august 2017)
5. Saqr, H. E., Pearl, D. K., & Yates, A. J. (1993). A Review and Predictive Models of Gang ioside Uptake by Biological Membranes. Journal of neurochemistry, 61(2), 395-441.
6. Sánchez, B. B., Alcarria, R., De Rivera, D. S., & Sánchez-Picot, A. (2016). Predictive algorithms for mobility and device lifecycle management in Cyber-Physical Systems. EURASIP Journal on Wireless Communications and Networking, 2016(1), 228.

7. Quartieri, J., Mastorakis, N. E., Iannone, G., Guarnaccia, C., D'ambrosio, S., Troisi, A., & Lenza, T. L. L. (2009, December). A review of traffic noise predictive models. In Recent Advances in Applied and Theoretical Mechanics, 5th WSEAS International Conference on Applied and Theoretical Mechanics (MECHANICS'09) Puerto De La Cruz, Tenerife, Canary Islands, Spain December (pp. 14-16).

8. Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A., & Wang, Y. (2003). Review of road traffic control strategies. Proceedings of the IEEE, 91(12), 2043-2067.

9. Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. Journal of transportation engineering, 129(6), 664-672.

10. Malo, J. E., Suarez, F., & Diez, A. (2004). Can we mitigate animal–vehicle accidents using predictive models?. Journal of Applied Ecology, 41(4), 701-710.

11. Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. Human Factors, 49(3), 376-399.

12. Lin, L., Li, J., Chen, F., Ye, J., & Huai, J. P. (2017). Road Traffic Speed Prediction: A Probabilistic Model Fusing Multi-Source Data. IEEE Transactions on Knowledge and Data Engineering.

13. Li, L., You, S., Yang, C., Yan, B., Song, J., & Chen, Z. (2016). Driving-behavior-aware stochastic model predictive control for plug-in hybrid electric buses. Applied Energy, 162, 868-879.

14. Massera Filho, C., Terra, M. H., & Wolf, D. F. (2017). Safe Optimization of Highway Traffic With Robust Model Predictive Control-Based Cooperative Adaptive Cruise Control. IEEE Transactions on Intelligent Transportation Systems.

15. Bordel, B., Alcarria, R., Robles, T., & Martín, D. (2017). Cyber–physical systems: Extending pervasive sensing from control theory to the Internet of Things. Pervasive and Mobile Computing, 40, 156-184.

16. Amini, S., Gerostathopoulos, I., & Prehofer, C. (2017, June). Big data analytics architecture for real-time traffic control. In Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2017 5th IEEE International Conference on (pp. 710-715). IEEE.

17. Götte, C., Keller, M., Haß, C., & Bertram, T. (2016). Model Predictive Planning and Control Applied to Critical Traffic Situations. ATZ worldwide, 118(9), 64-69.

18. Suh, J., Yi, K., Jung, J., Lee, K., Chong, H., & Ko, B. (2016). Design and evaluation of a model predictive vehicle control algorithm for automated driving using a vehicle traffic simulator. Control Engineering Practice, 51, 92-107.

19. Carrier, M., Apparicio, P., & Séguin, A. M. (2016). Road traffic noise in Montreal and environmental equity: What is the situation for the most vulnerable population groups?. Journal of Transport Geography, 51, 1-8.

20. Wang, V. S., Lo, E. W., Liang, C. H., Chao, K. P., Bao, B. Y., & Chang, T. Y. (2016). Temporal and spatial variations in road traffic noise for different frequency components in metropolitan Taichung, Taiwan. Environmental Pollution, 219, 174-181.

21. Pappas, G. J., Lafferriere, G., & Sastry, S. (2000). Hierarchically consistent control systems. IEEE transactions on automatic control, 45(6), 1144-1160.

22. Dufour, L., Genoud, D., Ladevie, B., & Bezian, J. J. (2016, March). Heating and hot water industrial prediction system for residential district. In Advanced Information Networking and Applications Workshops (WAINA), 2016 30th International Conference on (pp. 821-826). IEEE.