Transportation Research Part C 38 (2014) 167–176

Contents lists available at ScienceDirect

Transportation Research Part C

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

# Using the Bayesian updating approach to improve the spatial and temporal transferability of real-time crash risk prediction models

Chengcheng Xu [a,*], Wei Wang [a], Pan Liu [a], Rui Guo [b], Zhibin Li [a]

[a] Jiangsu Key Laboratory of Urban ITS, Southeast University, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Si Pai Lou #2, Nanjing 210096, China

[b] Department of Civil and Environmental Engineering, University of South Florida, 4202 E. Fowler Ave., Tampa, FL 33620, United States

ARTICLE INFO

*Article history:*
Received 7 September 2013
Received in revised form 16 November 2013
Accepted 18 November 2013

*Keywords:*
Real-time crash risk assessment
Spatial and temporal transferability
Bayesian updating
Freeway
Safety

ABSTRACT

This study aimed to improve the spatial and temporal transferability of the real-time crash risk prediction models by using the Bayesian updating approach. Data from California's I-880N freeway in 2002 and 2009 and the I-5N freeway in 2009 were used. The crash risk models for these three datasets are quite different from each other. The model parameters do not remain stable over time or space. The transferability evaluation results show that the crash risk models cannot be directly transferred across time and space. The updating results indicate that the Bayesian updating approach is effective in improving both spatial and temporal transferability even when new data are limited. The predictive performance of the updated model increases with an increase in the sample size of the new data. In addition, when limited new data are available, updating an existing model is better than developing a model using the limited new data.



## 1. Introduction

The development of advance traffic management systems (ATMS) has recently prompted research into proactive safety management strategies to prevent crashes on freeways (Lee et al., 2006; Abdel-Aty et al., 2007; Allaby et al., 2007). Considerable research has been conducted to develop real-time crash risk prediction models using traffic data collected from freeway surveillance systems (Oh et al., 2001; Lee et al., 2003; Abdel-Aty et al., 2004, 2012; Ahmed and Abdel-Aty, 2013; Hossain and Muromachi, 2011; Xu et al., 2012, 2013a; Golob et al., 2008; Pande et al., 2011). These models reveal the relationship between crash risks and traffic flow characteristics, which is critical to the development of proactive safety management strategies on freeways. They can help identify hazardous traffic patterns with high crash potential, where proactive safety management strategies would be used to reduce the crash risks by smoothing the traffic flow (Lee et al., 2006; Abdel-Aty et al., 2007). Accurate and reliable real-time crash risk models are needed for the real-time operations of proactive traffic management systems to improve freeway safety. Thus, it is desirable to obtain reasonably accurate models that have good predictive performance and good transferability across time and space.

Previous studies suggested that real-time crash risk models were effective in identifying hazardous traffic conditions that may lead to crashes (Abdel-Aty et al., 2004, 2012; Xu et al., 2012, 2013a, 2013b; Golob et al., 2008; Pande et al., 2011; Ahmed

* Corresponding author. Tel.: +86 13801580045.
*E-mail addresses:* iamxcc1@gmail.com (C. Xu), wangwei@seu.edu.cn (W. Wang), liupan@seu.edu.cn (P. Liu), rui@mail.usf.edu (R. Guo), lizhibin-2002@163.com (Z. Li).

0968-090X/$ - see front matter 
http://dx.doi.org/10.1016/j.trc.2013.11.020

et al., 2012). However, relatively few studies have focused on the transferability of real-time crash risk models. Pande et al. (2011) examined the spatial transferability of the crash risk model, which was developed on a freeway corridor, and then applied to other freeway corridors. The results showed that the developed model produced very poor predictive performance for other freeway corridors with different traffic patterns. In a subsequent study, Shew et al. (2013) developed a crash risk model based on the data collected from the US-101 freeway in California. The predictive performance of this model on the data collected from the nearby I-880 freeway was tested. The results showed that the difference in crash prediction accuracy between the US-101 dataset and the I-880 dataset was as large as 15% for the given false alarm rate. These results suggested that the real-time crash risk models could not be directly transferred from one freeway to another due to the differences in traffic flow conditions, driver populations, and the structure of the freeway surveillance systems. Thus, how to improve the transferability of real-time crash risk models is still a critical issue that needs to be addressed.

The quality and quantity of the high-resolution traffic data used to calibrate the real-time crash risk model is very important to the success or failure of these models. Models developed with limited data will not produce adequate prediction accuracy. The traffic data used to develop these models are usually collected from freeway surveillance equipment for long periods of time, such as one or two years. In addition, matching data from different sources and combining them is also a considerable effort (Shew et al., 2013). Thus, once a model is well-specified to capture the relationship between crash risk and traffic flow conditions for a freeway, it is cost-effective to transfer such a model to other freeways. The transferability of a crash risk model includes two aspects, spatial and temporal transferability. Spatial transferability involves the application of a crash risk model developed on a freeway for prediction on a different freeway for the same time period. Temporal transferability is using a crash risk model developed during one time period for predicting crash likelihood during another time period on the same freeway.

Several studies have been conducted to improve the transferability of aggregate accident prediction models (Hadayeghi et al., 2006; Sawalha and Sayed, 2006; Khondakar et al., 2010; Elvik, 2013; Chen et al., 2011). Hadayeghi et al. (2006) examined the temporal transferability of zonal-level accident prediction models. It was found that the model developed using 1996 data could not be used for predicting crash frequency in 2001. The Bayesian updating approach and calibration factors were used to update the 1996 models. The updated models could produce reasonably good predictive performance on the 2001 sample. Sawalha and Sayed (2006) compared three different methods for improving the transferability of the negative binomial models. A maximum likelihood method was proposed for recalibrating the transferred model. The results showed that the maximum likelihood method was better than the method proposed by the Interactive Highway Safety Design Model (IHSDM). Chen et al. (2011) applied the Bayesian model averaging (BMA) approach to improve the transferability of aggregate accident prediction models. The results showed that the BMA approach was superior to conventional model calibration methods. Most of these studies focused on the transferability of aggregate crash prediction models. However, similar research conducted to improve the transferability of the real-time crash risk models could not be found.

This study aimed to use the Bayesian updating approach to improve the spatial and temporal transferability of the real-time crash risk prediction models. This study has two basic objectives. First, it examines the spatial and temporal transferability to assess whether the relationship between crash likelihood and real-time traffic flow variables holds reasonably well across time and space. Although previous studies suggested that the crash risk models developed for one freeway could not be directly transferred to another freeway due to the difference in traffic patterns (Pande et al., 2011), the temporal transferability of the crash risk models is still not addressed. The second objective is to use the Bayesian updating approach to improve model transferability. This paper also examines the effects of the size of the new data used for updating on the model transferability.

## 2. Data sources

The data were collected from a 17-mile section on the I-880N freeway and a 20-mile section on the I-5N freeway in California, United States. There are 36 and 33 loop detector stations along the two selected freeway sections, respectively. Three data samples were collected: one from the selected I-880N freeway section in 2002, one from the selected I-880N freeway section in 2009; and one from the selected I-5N freeway section in 2009. The crash data were obtained from the Statewide Integrated Traffic Records System (SWITRS) maintained by the California Department of Transportation (Caltrans). The numbers of identified crashes in the three samples were 276, 508, and 458, respectively. The real-time traffic data were obtained from the Highway Performance Measurement System (PeMS), which is also maintained by the Caltrans.

The PeMS database provided 30-s raw loop detector data, including vehicle count, vehicle speed, and detector occupancy. For each crash in the data samples, traffic data were extracted in the time interval between 5 and 10 min prior to crash occurrence. For example, if a crash occurred at 8:15 PM, the considered traffic data were collected from 8:05 to 8:10 PM. The purpose was to compensate for any inaccuracies in the reported crash occurrence time (Christoforou et al., 2010; Golob and Recker, 2004). Doing so also helps to identify hazardous traffic conditions ahead of crash occurrence time to make preemptive measures possible (Pande et al., 2011; Xu et al., 2013c). The 30-s raw detector data collected from two consecutive upstream–downstream detector stations were aggregated into 5-min intervals and converted into the 23 traffic variables presented in Table 1. Due to the high correlation between traffic variables on adjacent lanes, these traffic variables were aggregated over all lanes. The averages and standard deviations were obtained by using the traffic values combined across