Development of Real-time Crash Prediction Model with Simulated Detectors

Ananya ROY[1], Yasunori MUROMACHI [2], Moinul HOSSAIN[3]

[1]Graduate Student, Graduate School of Inter Disciplinary Science and Engineering, Tokyo Institute of Technology
(Nagatsuta 4259, Yokohama, Kanagawa, 226-8502 Japan)
E-mail:roy.a.aa@m.titech.ac.jp
[2]JSCE member 198830920, Associate Professor, School of environment and Society, Tokyo Institute of Technology
(Nagatsuta 4259, Yokohama, Kanagawa, 226-8502 Japan)
E-mail: muromachi.y.aa@m.titech.ac.jp
[3]Assistant Professor, Islamic University of Technology
(Board Bazar, Gazipur, 1704, Bangladesh)
E-mail: moinul@iut-dhaka.edu; moinul048i@yahoo.com

Several Real-Time Crash Prediction models have been built as a tool to augment road safety. Hence arises the question of the quality and the dissemination of the available traffic data obtained from traffic detectors. Traffic state estimations and predictions are often inaccurate due to inaccurate detector locations. Several studies have shown the influence of detector locations on traffic state estimation. The motivation for this research is to find a way to deploy finite resources and generate an efficient traffic detection system. In this study, we selected a 1.40 km long segment with a three-legged junction of Route 3 Shibuya Line of Tokyo Metropolitan Expressway to show how the evolution of traffic flows can be predicted over time based on simple macroscopic model. This segment is modelled by a set of cells which tracks traffic state over time with generalized Cell Transmission Model (CTM) which is consistent with kinematic wave theory of traffic flow. Traffic data generated from both fixed detectors and CTMs were used to build Real-time Crash Prediction models (RTCP) based on Bayesian Network- a probabilistic method. Performance of both methods are compared. The results demonstrate that the proposed CTM method can predict traffic conditions with an accuracy of 90% when the traffic data is collected from 300 meters up and downstream of the crash location. This accuracy can be improved further by introducing more traffic flow parameters like velocity and using more traffic data.

Key Words : *Bayesian Network, Cell Transmission Model, Real-Time Crash Prediction*

1. INTRODUCTION

Over the last decade, the idea of predicting crash in real-time as part of active traffic management (ATM) has attracted substantial attention among researchers in the field of road safety. The concept of real time crash prediction is based on the hypothesis that the probability of a crash on a specific road section can be predicted for a very short time window using the instantaneous traffic flow data[3,4,5,6,7,8)]. Substantial effort has been put in improving the Real-time Crash Prediction models (RTCP) [1,2)]. Hence, data collection system plays a significant role in the performance of the RTCP models[11,12)].

Moreover, with the advancement of Intelligent Transportation System (ITS) and development of advanced transportation information systems (ATIS), numerous traffic data collection systems have arised[13,14,15)]. Accurate and reliable estimate of real-time traffic data is essential for optimizing network performance during unpredictable traffic incident.

Traditionally, traffic flow models have been used to predict traffic state on segments of freeway or urban networks. Increase in number of traffic sensors and availability of real-time traffic measurements have facilitated utilization of measured traffic variables to improve overall estimate of traffic state. Inspite of that, these sensors still do not cover every part of the network and cannot provide a complete picture of existing traffic state with higher spatial resolution. Moreover, there are several situations in reality where the fixed sensors are installed far away from each other with variable spacings. This could result in different outputs from various traffic models.

Furthermore, it becomes difficult to compare the traffic models due to a variety of detector layout for individual cases or road segments. In traffic state estimation prediction from traffic flow models, which can be of higher spatial resolution is combined with real-time measurements to get a final estimate with higher spatial and temporal resolution.Thus, real-time traffic state estimation refers to estimation of traffic flow variables (traffic flow, density) for a segment of road or network with an adequate time and space resolution based on limited available measurements from traffic sensors[16].

Previous studies employed comprehensive methodology of estimating traffic state using real-time traffic data from sensors and prediction of traffic state from a second order traffic flow model. In this estimation model, parameters of second order traffic flow models such as free-flow speed and critical density were converted into stochastic variables by using random-walk equations and estimated for each time-step[17,18]. Another study proposed a framework that utilized particle filtering algorithm with second order traffic flow model and unscented Kalman filter algorithm with macroscopic traffic flow model to estimate traffic for a freeway section[19,20,21]. In addition, a Cell Transmission Model (CTM) based Switch Mode Model (SMM) was derived based on five different traffic modes to avoid non-linearity caused due to nature of fundamental traffic flow diagram in CTM [22,23]. Later on, a study showed effect of sensor location when there are constrains on the number of sensors which can be installed over a network[12]. They used CTM with ensemble Kalman filter (EnKF) to study the impact of various sensor location configurations on estimation of travel speed and concluded that sensors located at large distances from each other without location optimization lead to overestimation of travel speed, whereas sensor numbers can be reduced if their locations are optimal to achieve a better estimate of travel speed. Many other studies have focused on optimization of sensor location to find the minimum number of traffic sensors to cover a road network [13,14,24,25]. Thus, it is evident from previous studies that location of traffic sensors or detectors influences the estimationof traffic state hence estimation of traffic flow data leading to influencing the performance of the RTCP models. Crashes occurring on freeways/expressways are considered to relate closely to immediate traffic conditions occurred before the crash, which are time-varying.

In this study, sensors or detectors are located approximately 250 meters apart from each other in Tokyo Metropolitan Expressway's Route 3 Shibuya Line. A general CTM model was produced for the most crash prone location of the route. The objective of this study is to investigate the performance of the RTCM based on Bayesian Network (BN) while the traffic data is collected from uniformly and densely distributed sensors

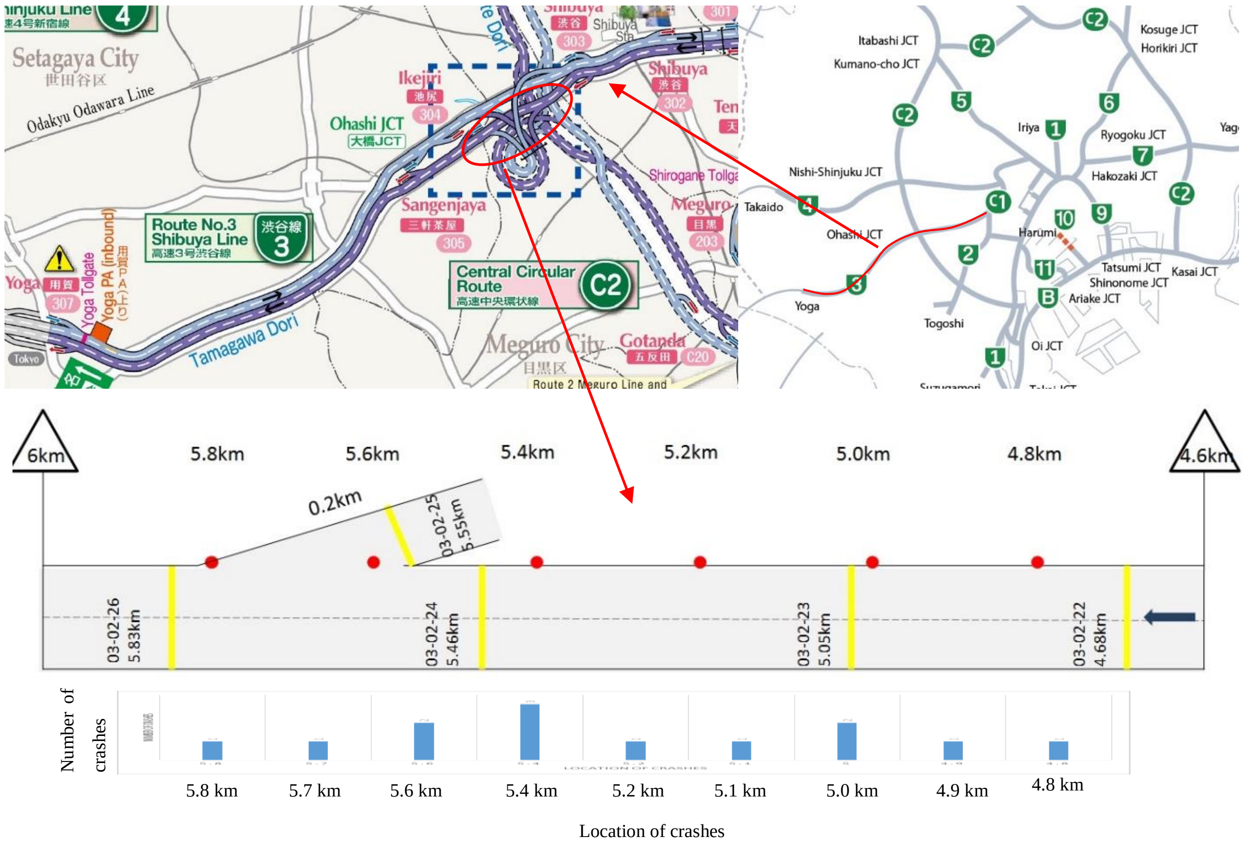

Fig.1 Route 3 Shibuya Line, Tokyo Metropolitan Expressway, Japan
Source: Official Website of Tokyo Metropolitan Expressway Company Limited

or detector[9,10,26,27].

The paper is organized in five sections. This section has already addressed the problems with the present crash prediction models and the necessity of detector locations. In the second section, the data collection is discussed. The third section discusses the methodology of this study and the forth section deals with the model building and validation. Lastly, the fifth section draws conclusion and future scope of this study.

2. DATA

(1) Study Area

Since the suitable study area for this study requires to have sufficient crash cases and closely spaced detectors, Route 3 Shibuya Line of the Tokyo Metropolitan Expressway was selected. It is about 11.9km long and one of the busiest expressways in Japan. It is connected with the Tomei Expressway starting at the point on the boundary of the Tokyo Metropolitan Ward Area as shown in **Fig. 1**. It sustains a large number of crashes throughout the year as well. It is a radial route connected with the inner circular loop of Tokyo Metropolitan Expressway, widely known as C1, which serves the central part of Tokyo. The expressway has two lanes in each directions with 48 detectors in the inbound direction and 46 detectors in the outbound direction (about 250 meters apart). A total of 310 crash cases were reported for the route out of which 100 crash cases were found to be executable for modeling purpose after eliminating the errorneous and missing data.. For this study we chose a 1.40 km long segment between kilopost 4.6km and 6.0km in the outbound direction since it is the most crash prone (thirteen crash cases in sixe months) location. The distribution of crashes in this segment of the route is shown in **Fig. 1** and the list of crash cases are shown in **Table 1**.

Table-1: List of crashes in the segment of study

Date of crash	Time of Crash	Location of crash (km)	Up-stream detector	Down-stream detector
2014-03-05	14:34	5.1	03-02-23	03-02-24
2014-03-07	12:48	5.4	03-02-23	03-02-24
2014-03-07	21:55	5	03-02-22	03-02-23
2014-03-13	18:17	4.9	03-02-22	03-02-23
2014-04-02	08:40	5	03-02-23	03-02-24
2014-04-13	11:02	5.7	03-02-24	03-02-26
2014-04-22	17:35	5.4	03-02-23	03-02-24
2014-05-04	17:22	5.6	03-02-24	03-02-26
2014-05-16	20:31	5.8	03-02-24	03-02-26
2014-07-18	10:59	4.8	03-02-22	03-02-23
2014-08-02	07:07	5.4	03-02-23	03-02-24
2014-08-17	22:18	5.2	03-02-23	03-02-24
2014-08-19	13:52	5.6	03-02-24	03-02-26

(2) Data Collection and Processing

The study investigates the crash mechanism with real-time traffic data. Hence, the traffic flow database has been collected from loop detector data between March and August 2014.The detectors in the study area store data of speed, vehicle count, occupancy and number of heavy vehicles for 24 hours a day, 365 days a year. Six months detector data and crash data were extracted. Detector data consists of detector location (kilo post), speed (1min average speed in km/hr), flow (1 min aggregated in veh/min) and occupancy (%). Crash data contained information about date, time (in minutes), location (to nearest 10 meters), crash lane, type of crash and vehicle involvement.

For the selection of high performance detectors, we investigated the accuracy of detectors based on missing or errorneous data. The accuracy of the detectors of the study segment are listed in **Table 2**. Traffic data were divided into two categories- (a) crash data and (b) normal data for the models. For example, if a crash occurred on 3rd March, Wednesday at 15:00, then the crash and normal data were collected in the manner shown in **Table 3**.To avoid misleading data, we removed all normal condition data where a crash took place on the same date before or after 1 hour of the selected time period. Also, crash data with any

Table-2: List of detectors in segment of study

Detector Number	Accuracy (%)	Detector Number	Accuracy (%)
03-02-22-1	97.2	03-02-22-2	97.7
03-02-23-1	99.3	03-02-23-2	99.2
03-02-24-1	99.4	03-02-24-2	99.4
03-02-25-1	99.3	03-02-25-2	99.3
03-02-26-1	99.4	03-02-26-2	99.4

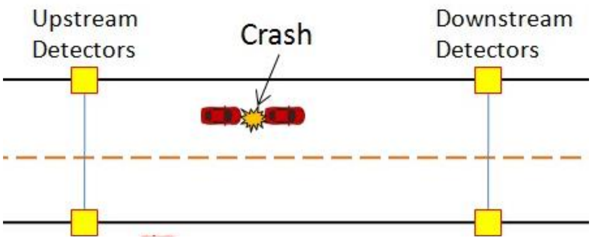

Fig. 2 Layout of expressway and data collection procedure.

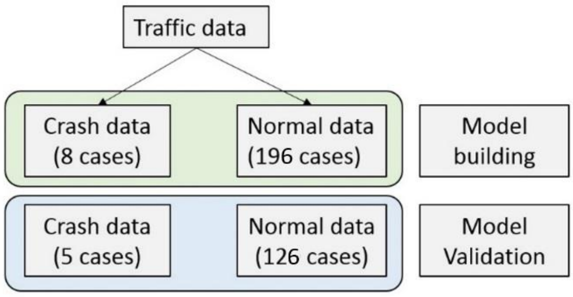

Fig. 3 Data management for RTCP model

missing data was ignored. After all these screening, for this study a total of 13 numbers of crash and 312 numbers of normal data left for model building. Out of 13 crash cases, 8 cases and corresponding 192 normal cases were used for model validation (**Fig. 3**). In order to develop a real-time crash prediction model

Pairs of detectors- nearest upstream and nearest downstream of the crash location was considered for the entire route of the expressway[8]. The data collection arrangement is as illustrated in **Fig. 2**.

For cell transmission model (CTM), data is prepared separately. There are three basic parameters for constructing a CTM, i.e. maximum flow or, capacity (veh/hr), jam density (veh/km) and free-flow speed (km/hr). All these three parameters are calculated from the fundamental diagram (FD) of traffic flow. An example is given in **Fig. 4**. The entire segment is divided into four arcs (0 to 3) and five nodes (0 to 4) as shown in **Fig. 5**. The FDs were produced for each arc and for every single day (335 days in total). The methodology of BN and CTM are explained in the next section.

Table 3 Crash data and Normal data collection procedure.

Crash Occurrence Time	Crash Data	Normal Data
3rd March, Wednesday at 15:00	3rd March, 14:45 to 15:00	All other Wednesdays in March, 14:45 to 15:00

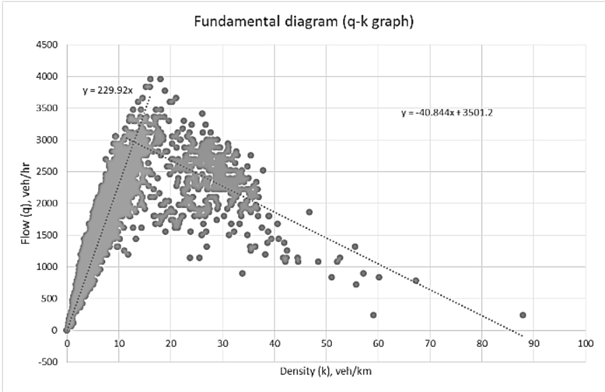

Fig. 4 Fundamental traffic flow-density diagram

There are two types of variables in Bayesian Network- information variable and hypothesis variable. Information variables are those, the values of which are to be expected to be obtained to calculate the probability of the hypothesis variable. For the purpose of model building, four absolute values: upstream flow(u_q), downstream flow(d_q), upstream occupancy(u_o), downstream occupancy(d_o) along with two relative values of traffic variables i.e. upstream and downstream difference of flow (diff_q) and occupancy (diff_o) were used as information variables. Since we have employed the generalized CTM, velocity data were not used for this study.

Bayesian Network, is a probabilistic graphical modeling method where we represent a system with a graph and a joint probability distribution compacted with the notion of conditional independence. Later, we can use this model of system to understand the dynamics within the system and also to predict the state of variables in lights of the evidence on any one or more variables.

3. METHODOLOGY

(1) Bayesian Network Model

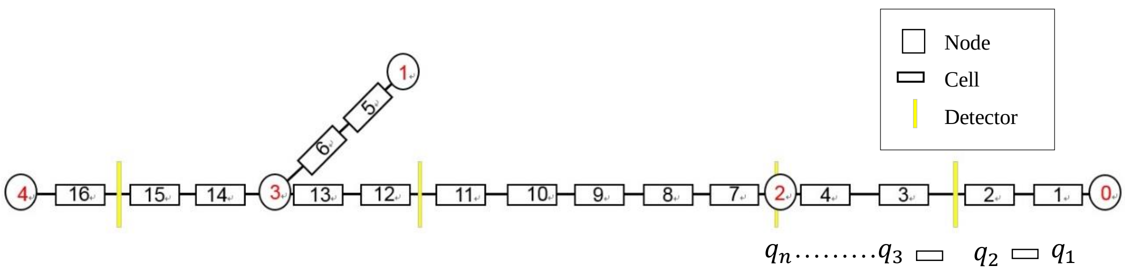

Fig. 5 Layout of the segment for cell transmission model

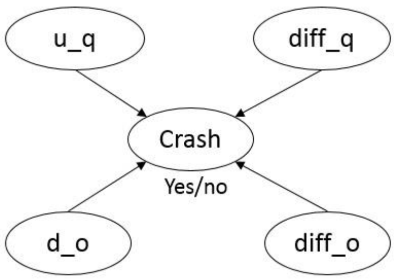

Fig. 6 Structure of a sampleModel built with Bayesian Network

Fig. 6 presents a simple BN involving four variables. Here, each variable is represented with a node and the influence of one variable on others is demonstrated with directed edges (may or may not represent causality). We would like to mention here that these graphs are acyclic in nature and are called acyclic directed graph (DAG).Hence, the BBN in **Fig. 6** can be written as:

P(*crash,u_q,diff_q,d_o,diff_o*)=P(*diff_o*).P(*d_o|diff_o*).P(*diff_q|d_o,diff_o*).P(*u_q|diff_q,d_o,diff_o*).P(*crash | u_q,diff_q,d_o,diff_o*) (1)

The task of EM-Algorithm (expectation maximization) in BBN is to deter-mine the conditional probability tables (CPTs) for nodes based on prior probabilities and availability of new N number of records. The algorithm has two steps – calculating the expected sufficient statistics and then maximizing its likelihood. To elaborate more, if no probability is assigned to a variable for which we are estimating the parameters, a uniform distribution is assumed. Then, with presence of a batch of data, the new parameter is estimated in such way that first, the expected sufficient statistics under that parameter is calculated and then the log-likelihood of that parameter under the expected sufficient statistics is maximized. This is an iterative process and it stops when one of these two criteria are satisfied – i) the maximum number of iteration specified by the user has exceeded, or, ii) the relative log-likelihood between two succes-sive iterations is smaller than the preset minimum dif-ference value. Here, it is important to mention that the EM-algorithm does not need data on each of the vari-ables to update the model. The adaptation algorithm is similar to the EM-algorithm with the exception that here the evidence from each record is propagated throughout the network and the parameters for each of the variables are updated accordingly.

The BN based RTCP model has been built and evaluated in three interlinked stages. While the first stage selects variables, second stage constructs the model by finalizing the directed acyclic graph (DAG) and then generates the conditional probability tables for each variable. The last stage evaluates the model performance using a separate dataset which lays out a strategy for implementing the model in real-life situations.

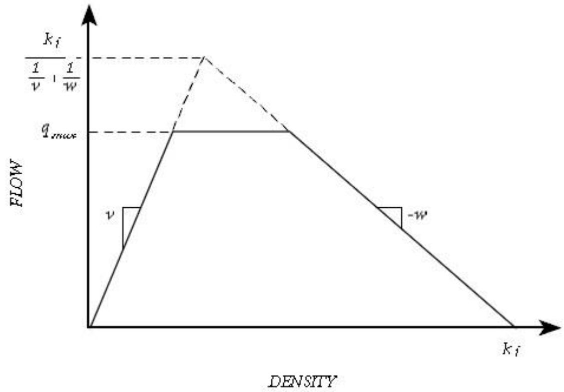

Fig. 7 The equation of state of the cell transmission model

(2) Cell Transmission Model

In Daganzo[26,27,29)] it is shown that if the relationship between traffic flow (q) and density (k) is of the form depicted in Fig. 7:

q = min { vk , q_{max}, w (k_j-k) }, for 0≤ k ≤ k_j, (2)

then the LWR equations for a single highway link can be approximated by a set of difference equations where current conditions (the state of the system) are updated with the tick of a clock. In the above expression v, qmax, w and kj are constants denoting respectively: the free-flow speed, the maximum flow (or capacity), the speed with which disturbances propagate backward when traffic is congested (the backward wave speed), and the maximum (or jam) density.

The method assumes that the road has been divided into homogeneous sections (cells), i, whose lengths equal the distance traveled by free-flowing traffic in one clock interval. (Although a closer approximation to the LWR results is obtained with short cell lengths (e.g. 100 meters) the procedure can be applied with cells of any length.) The state of the system at instant t is then given by the number of vehicles contained in each cell, ni(t). The following parameters are defined for each cell:

Ni(t), the maximum number of vehicles that can be present in cell i at time t,and

Qi(t), the maximum number of vehicles that can flow into cell i when the clock advances from t to t+1.

These constants can vary with time (e.g. as per the occurrence of transient traffic incidents) but this dependence will be ignored in this paper for simplicity of notation. The first constant is defined to be the product of the cell’s length and its jam density, and

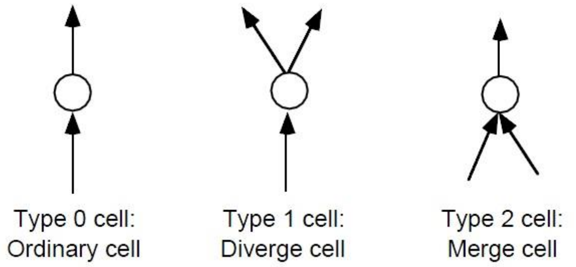

Fig. 8 Cell type specifications

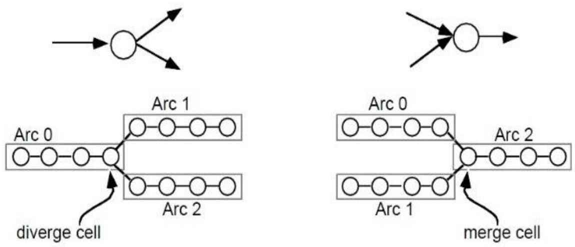

Fig. 9 Merge and diverge cells in a junction

the second one the product of the clock interval and the cell's capacity. If cells are numbered consecutively starting with the upstream end of the road from i = 1 to I, the recursive relationship of the cell-transmission model can be expressed as:

$$n_i(t+1) = n_i(t) + y_i(t) - y_{i+1}(t) \quad (3)$$

where y_i (t) is the inflow to cell i in the time interval (t,t+1), given by:

$$y_i(t) = \min\{\, n_{i-1}(t),\ Q_i(t),\ d\,[N_i(t) - n_i(t)] \,\} \quad (4)$$

where d= w/v. Note the similarity of (2) and (4).

For finite roads, boundary conditions can be specified by means of input and output cells. The output cell, a sink for all exiting traffic, should have infinite size (N_i+1 = ∞) and a suitable, possibly time-varying, capacity. Input flows can be modeled by a cell pair. A "source" cell numbered "00" with an infinite number of vehicles (n_{00} (0) = ∞) that discharges into an empty "gate" cell "0" of infinite size, N_0 (t) = ∞. The inflow capacity Q_0 (t) of the gate cell should then be set equal to the desired link input flow for the corresponding time interval. The gate cell then acts as a metering device that releases traffic at the desired rate while holding (as a parking lot would) any flow that is unable to enter the link. (Although it may be possible to eliminate gate cells in an efficient computer implementation, the program logic should preserve their effects. Gate cells ensure that any time dependent O-D table can be handled; i.e. that the LWR problem is well-posed for all O-D tables.)

Network representation

A general transportation network is usually described by a directed graph of nodes and arcs, including some physical data for each arc. An arc's physical data includes its length and the parameters defining a q-k relation for a steady state of traffic of the type shown in **Fig. 7**.

It should be clear that each arc of a network must have an appropriate boundary conditions at both of its ends have been defined. Thus, we assume that each arc of the graph has been subdivided into cells, as explained above. There are three kinds of cells- 1) ordinary, 2) diverge and, 3) merge cell as shown in **Fig. 8.** And the cells are organized inside the arcs. A road section can be divided into several arcs for CTM (**Fig. 9**).

The general procedure for networks involves two steps for each tick of the clock:

(i) Determine the flow on each link with the equivalent of equation (4).

(ii) Update the cell occupancies by transferring the flows of step (i) from the beginning-cell to the end-cell of each link.

4. MODEL BUILDING AND VALIDATION

Table 4 RTCP models and corresponding information variables

Models	Information variables
Model-1	d_o, diff_o, diff_q, u_q
Model-2	u_o, diff_o, diff_q, d_q
Model-3	d_o, diff_o, diff_q, d_q
Model-4	u_o, diff_o, diff_q, u_q

Table 5 Arc information for CTM

Arc Nos.	U/s node	D/s node	Arc length, km	Free-flow speed, v_{ff} ($\frac{km}{hr}$)	Maximum flow, q_{max} (veh/hr)	Jam density, k_j,(veh/km)
,Arc 0	0	2	0.40	159	2100	413.77
Arc 1	1	3	0.20	196	2900	272
Arc 2	2	3	0.70	171	2800	203
Arc 3	3	4	0.30	188	3100	192

Table 6 confusion matrix for model-1 (method-1)

Model- 1 (average 24%)		BN		BN (CTM)	
		Actual		Actual	
		crash	no crash	crash	no crash
Pre-dicted	crash	14	80	10	113
	no crash	16	462	20	429
overall accu-racy, %		83.22		76.75	

Table 7 confusion matrix for model-2 (method-1)

Model- 2 (average 29%)		BN		BN (CTM)	
		Actual		Actual	
		crash	no crash	crash	no crash
Pre-dicted	crash	6	119	5	180
	no crash	69	1771	70	1710
overall accu-racy, %		90.4		87.3	

(3) Model Building

For BN based RTCP model building, six information variables were chosen: upstream flow(u_q), downstream flow(d_q), upstream occupancy(u_o), downstream occupancy(d_o), upstream and downstream difference of flow (diff_q) and occupancy (diff_o). Using different combinations of these six information variables four BN models were generated (**Table 4**). Since the study segment is not long enough, 15 minutes before the crash data were employed. Data was collected from the nearest upstream and down stream detectors as mentioned earlier (**Fig. 2**). All the BN models were created with Hugin Expert A/S 8.2 software. And the CTM was produced using NETCELL software.

On the otherhand, in case of CTM, the entire length of the segment was divided into 4 arcs with varying lengths (**Table 5**) with a merging point at node 3 (**Fig. 5**). Each arc has its own traffic parameters i.e. q_{max}, v_{ff}, k_j for every day under this study. An example data is shown in **Table 5** for one day.

Table 8 confusion matrix for model-3 (method-1)

Model- 3 (average 27%)		BN		BN (CTM)	
		Actual		Actual	
		crash	no crash	crash	no crash
Pre-dicted	crash	6	113	5	195
	no crash	69	1777	70	1695
overall accu-racy, %		90.7		86.5	

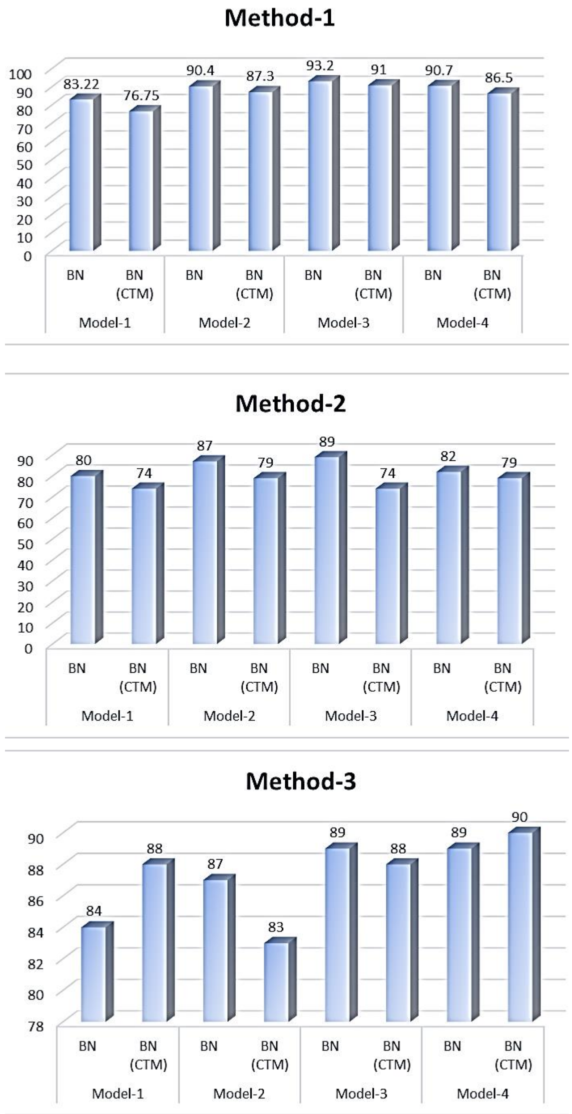

Fig. 11 Comparison of overall preformances of for BN and CTM models using three types of detector arrangements

For data collection, three methods were followed- 1) data collected from the upstream and downstream cells nearest to the crash location, 2) data collected from the upstream and downstream cells which are 2nd nearest to the crash location, and 3) data collected from the upstream and downstream cells which are 3rd nearest to the crash location. For example, in **Fig. 5,** for a crash that took place at 5.4 km

Table 9 confusion matrix for model-4 (method-1)

Model- 4 (average 25%)		BN		BN (CTM)	
		Actual		Actual	
		crash	no crash	crash	no crash
Pre-dicted	crash	7	66	9	111
	no crash	68	1824	66	1779
overall accu-racy, %		93.2		91.0	

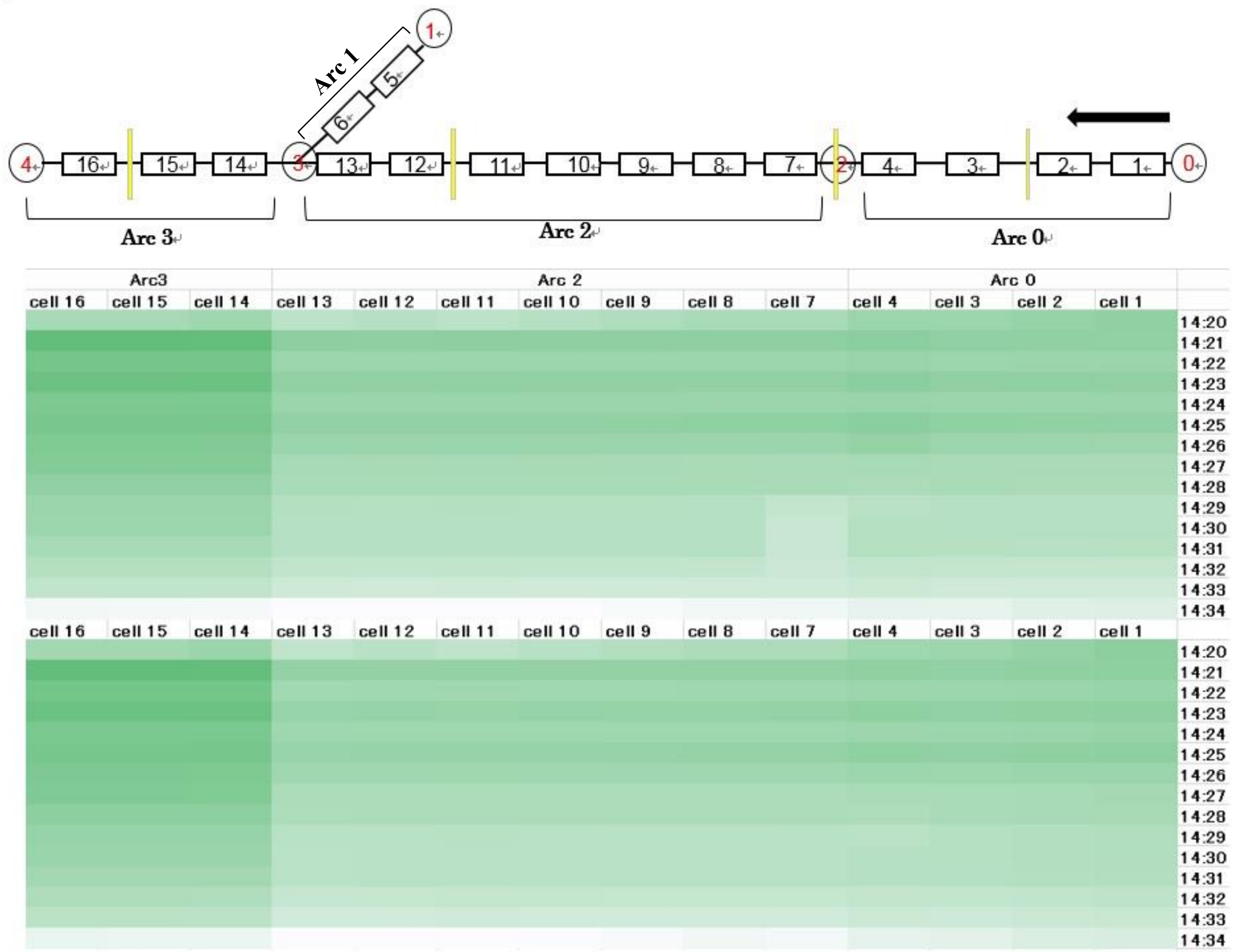

Fig. 10 Flow distribution within cells for a crash case (5th March, 2014; 14:34pm) and corresponding one no crash case using CTM

location, then for method 1, data will be collected from cell 11 and 12, for method 2, cell 10 and 13; for method 3, cell 9 and 14.

(2) Model Validation

As mentioned earlier, 8 crash cases and corresponding 196 normal cases were separated randomly for model validation. The success of the model depen d on its combined performance to predict crash and normal traffic conditions. After building BN models, validation data of crash and normal cases was entered individually in the models. Their associated probabilities to belong to crash prone traffic condition have been calculated based on the prior probabilities. If no new evidence is entered into the BN then the average probability of a traffic condition being associated with crash is 10.7% (BN model-1). This value has been used as the minimum threshold for evaluating the performance of the model. Also, a false alarm rate 6.9% is considered to check the prediction capability of the BN models. Threshold value higher than the average probability is also checked for comparison purpose. The confusion matrices are s hown in **Table 6** to **9** for comparison.

The equations employed for the calculations are:

*Crash= (Calculated probability over threshold/Crash sample size)*100* (5)

*Normal= (Calculated probability below threshold/Normal sample size)*100* (6)

*Overall= (Total correct classification/Total sample size)*100* (7)

It is understandable from the traffic flow phenomena, that a that a sudden change in flow i.e. increase or decrease in flow could be a vital factor to explain crash likelihood. Sudden increase in flow could happen due to quick dissipation of queues causing a backward shock wave. On the contrary, a disruption in the downstream that propagates a shock wave to the upstream might be responsible for a sudden decrease of flow. There have been many cases where a lower upstream flow followed by a higher downstream flow resulting in a hazardous situation. Specially, during a situation when there is a sudden increase of upstream flow and a sudden decrease of downstream flow.

The crash prediction and overall prediction performance under method 2 and 3 are shown in **Fig. 11**. From the overall performances, it can be seen that general BN models performs better than the BN models build with traffic data obtained using CTM for the most of the cases. One reason for this is very close proximity of the cells we choose for method-1 (immediate up and down stream cell to the crash location). The variation of flow is less between these two cells. During 15 minute the change of flow between upstream and downstream was insignificant. But from **Fig. 11**, it can be seen that for method- 3, the overall performance (means joint performance of predicting crash likelihood and no-crash likelihood) model-1 and model-4 for CTM based BN model (88% and 90%) is better than general BN models (84% and 89%). The reason behind this might be the the distance of the traffic data collection point (or cells) from the location where crash incident took place. Previous studies have shown[12)] that the interdtector spacing more than 400 meters might result in a huge variation in traffic state estimation. In this study, the length of each cells were approximately 100 meters. Thus for method-3, traffic data were used from 3rd cell upstream and downstream fromt he incident location, which are about 300 meters from the crash location. A sample flow distribution diagram produced using CTM in a space-time variable space for a crash occurred on 5th March, 2014 at 14:34 pm is shown in **Fig. 10**. The horizontal axis represents the length of the study segment of the expressway where the traffic flow direction is right to left. Vertical axis is reflecting the 15 minute time range before the crash. In this figure, darker region means higher traffic flow and vice versa. The crash took place around 14:34 pm at 5.10km location and the nearest cells are number 4 and 7 **(Fig. 10**). The upper diagram is the record of the day of crash whereas the bottom diagram is for one corresponding no-crash day. From both diagrams, it is clear that the traffic flow has reduced with time and at the same time, flow has intensified from cell number 14 (in Arc 3). This makes sense because cell 14 is the merging cell of arc 1 and arc 2. Other than this, a slight drop of flow can be perceived in cell 7 at time14:29pm which continued until 14:34pm. Altough, the variation is not so prominent, there is a mild rise in flow in cell 4, which is upstream of the crash location. On the contrary, the diagram for no-crash day seems have no variation of flow along the length of the segment. Hence, the CTM is able to capture the traffic flow data in cells from the data collected from the fixed detectors.

5. CONCLUSIONS

The focus of this study was to investigate if cell transmission model (CTM) could be applied for producing uniformly spaced detectors while traffic data from only fixed detectors are present. In addition, to compare the performance of the Bayesian Network based RTCP model generated with the data from fixed detectors with variable spacings and the data generated from CTM. And it can be seen from the retults that CTM can produce traffic flow data from the fixed detector data. Privous studies showed that depending on the inter-detector spacing, the quality of the data could vary leading to a very different traffic state estimation. In this study data from three different detector spacings were investigated. Altough, the outcome of the CTMs showed no prominent evidence of a better performance compared to general BN based models, it was observed that the CTM can be employed for traffic data generation. The possible reasons could be lack of proper traffic flow data i.e. velocity, congestion index etc. inaddition to the flow and occupancy data. However, there are several scopes of improvement for these models. For better functioning of the models, velocity based CTM-v could be a better choice. Previous studies showed that velocity has a prominent impact on the development of RTCP models. This study did not use velocity as an information variable since generalized CTM can not be used for such calculation. Furthermore, a filtering method such as Kalman Filter, EnKF or Particle Filter etc. methods could be applied for more refined traffic state estimation. Moreover, for the benefit of simplicity, only one segment of the expressway was selected for this study which contains only thirteen crash cases. Performance of a real-time crash prediction model largely depends on the quality as well as the quantity of the data. Another probably reason behind the uncertain results could be less study concluding the influence of inter-detector spacing in traffic flow estimate phenomena. There are several studies on this subject, but very studies have specific claims of the optimum position of the detector locations for traffic data collection or traffic state estimation.

Hence there are many scopes of improvement in the future. For example, introduction of velocity based CTM for generating more traffic flow parameters. Also, to investigate how to extend this model for a larger segments with several inflow-outflow ramps. In addition, since this CTM is calculation intensive model, it requires a long time to analyze one small segment, thus transferability of the model to other segments can be investigated as well.

ACKNOWLEDGMENT: We would like to thank Tokyo Metropolitan Expressway Company Limited for providing necessary data to conduct this study. This study is also supported by JSPS KAKENHI Grant Number 25420535.

.

REFERENCES

1) Oh, C., Oh, J., Ritchie, S., Chang, M.:Real-time estimation of freeway accident likelihood. *Presented at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C,* 2001.
2) Lee, C., Saccomanno, F., Hellinga, B.: Real-time crash prediction model for the application to crash prevention in freeway traffic. *Transp. Res. Rec. 1840, 67–77,* 2003.
3) Lee, C., Szccomanno, F., and Hellinga, B.:Analysis of Crash Precursors on Instrumented Freeways. *In Transportation Research Record: Journal of the Transportation Research Board, No.1784, Transportation Research Board of the National Academics, Washington, D.C., pp. 1-8,* 2001.
4) Lee, C., Hellinga, B. and Saccomanno, F.: Proactive Freeway Crash Prevention Using Real-Time Traffic Control. Canadian Journal of Civil Engineering - *Special Issue on Innovations in Transportation Engineering, Vol. 30(6). pp. 1034-1041,* 2003.
5) Lee. C., Hellinga, B. and Saccomanno, F.: Real-Time Crash Prediction Model for the Application to Crash Prevention in Freeway Traffic. *In Transportation Research Record: Journal of the Transportation Research Board, No. 1840, Transportation Research Board of the National Academics, Washington, D.C., pp. 67-77,* 2003b.
6) Abdel-Aty M. and Pande A.:Identifying crash Propensity using Specific Traffic Speed Conditions, *Journal of Safety Research, Vol. 36 No.1, pp. 97-108,* 2005.
7) Hossain, M. and Muromachi, Y.: Development of a Real-time Crash Prediction Model for Urban Expressway, *Journal of the East Asian Society for Transportation Studies, Vol. 8, pp. 2092-2107,* 2010a.
8) Hossain, M. and Muromachi, Y.: Optimum Detector Spacing for Real-Time Monioring of Hazardous Locations on Urban Expressways. *Journal of Infrastructure Planning Review (JSCE), Vol27, pp. 1045-1054,* 2010b.
9) Hossain, M. and Muromachi, Y.: Applicability of bayesian network in real-time crash prediction, *Proceedings of Infrastructure planning review, Japan Society of Civil Engineers (JSCE), vol. 38*, 2008.
10) Hossain, M. and Muromachi, Y.: A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways, *Accident Analysis and Prevention 45 (2012) 373–381*, 2011.
11) Seo, T., Kusakabe, T. and Asakura, Y.: Traffic State Estimation with the Advanced Probe Vehicles using Data Assimilation, *2015 IEEE 18th International Conference on Intelligent Transportation Systems, 824-830, 2015.*
12) Hong, Z., and Fukuda, D.: Effects of Traffic Sensor Location on Traffic State Estimation, *15th meeting of the EURO Working Group on Transportation, Procedia - Social and Behavioral Sciences 54 (2012) 1186 – 1196, 2012.*
13) Eisenman,S.M., Fei, X., Zhou, X., and Mahmassani, N. S.: Number and Location of Sensors for Real-Time Network Traffic Estimation and Prediction Sensitivity Analysis, Transportation Research Record: *Journal of the Transportation Research Board, No. 1964, Transportation Research Board of the National Academies, Washington, D.C., pp. 253–259*, 2006.
14) Ahmeda,A., Watling, D., and Ngoduy, D.: Significance of sensor location in real-time traffic state estimation, *Fourth International Symposium on Infrastructure Engineering in Developing Countries,IEDC, Procedia Engineering 77 (2014) 114 – 122*, 2013
15) Viti, F., Verbeke, W., Tampère, C.M.J.,: Sensor locations for reliable travel time prediction and dynamic management of traffic networks, *Transportation Research Record: Journal of the Transportation Research Board, Vol. 2049, p. 103-110*, 2008.
16) Wang, Y., Papageorgiou, M.,Messmer, A.: Real-Time Freeway Traffic State Estimation Based on Extended Kalman Filter: Adaptive Capabilities and Real Data Testing.*Transportation Research Part A: Policy and Practice42. pp. 1340-58*, 2008.
17) Wang, Y., Papageorgiou, M.,: Real-Time Freeway Traffic State Estimation Based on Extended Kalman Filter: A General Approach. *[In English].Transportation Research Part B-Methodological 39, pp. 141-67*, 2005.
18) Papageorgiou, M., Blosseville,J. M., Hadjsalem, H.: Modeling and Real-Time Control of Traffic Flow on the Southern Part of Boulevard-Peripherique in Paris .1. *Modeling. [In English]. Transportation Research Part a-Policy and Practice 24, pp. 345-59*, 1990.
19) Ngoduy, D.: Kernel Smoothing Method Applicable to the Dynamic Calibration of Traffic Flow Models. *Computer Aided Civil and Infrastructure Engineering 26, pp. 420-32, 2*011.
20) Ngoduy, D.: Applicable Filtering Framework for Online Multiclass Freeway Network Estimation. *Physica A: Statistical Mechanics and its Applications387, pp. 599-616,* 2008.
21) Ngoduy,D.: Low-Rank Unscented Kalman Filter for Freeway Traffic Estimation Problems.*Transportation Research Record: Journal of the Transportation Research Board2260, pp. 113-22*, 2011.
22) Munoz, L., Sun,X. T., Horowitz, R., Alvarez, L.: Traffic Density Estimation with the Cell Transmission Model.*[In English].Proceedings of the 2003 American Control Conference, Vols 1-6, pp. 3750-55*, 2003.
23) Munoz, L., Sun,X. T., Horowitz, R., Alvarez, L.: Piecewise-Linearized Cell Transmission Model and Parameter Calibration Methodology.*Transportation Research Record: Journal of the Transportation Research Board 1965, 1, pp. 183-91*, 2006.
24) Lam, W. H. K., Lo, H. P.:Accuracy of Od Estimates from Traffic Counts.*Traffic engineering & control 31, pp. 358-67,* 1990.
25) Yang, H., Zhou, J.: Optimal Traffic Counting Locations for Origin-Destination Matrix Estimation.*Transportation Research Part B:Methodological 32, pp. 109-26,* 1998.
26) Daganzo, C. F.: The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory, *Transpn. Res.-B. Vol. 28B, No. 4, pp. 269-287,* 1994.
27) Daganzo, C. F.: The cell transmission model: network traffic, *California PATH Working Paper UCB-ITS-PWP-94-12*.
28) Cayford,R., Lin,W. and Daganzo, C. F.: The NETCELL simulation package: Technical description, *California PATH Research Report UCB-ITS-PRR-97-23*, 1997.

29) Daganzo C.F. :The cell-transmission model: A simple dynamic representation of highway traffic, *Transpn. Res.-B. Vol. 28B, No. 4, pp. 269-287*, 1993.

(Received April 27, 2017)