# INTEGRATED INCIDENT DECISION SUPPORT USING TRAFFIC SIMULATION AND DATA-DRIVEN MODELS

**Tao Wen, Corresponding Author**
Research Scientist, DATA61| CSIRO
13 Garden St, Eveleigh, 2015, NSW, Australia
Tel: (+61) 02 9490 4918; Email: tao.wen@data61.csiro.au

**Adriana-Simona Mihăiţă**
Research Scientist, DATA61|CSIRO
13 Garden St, Level 5, Eveleigh, 2015, NSW
Email: simona.mihaita@data61.csiro.au

**Hoang Nguyen**
Research Scientist, DATA61|CSIRO
13 Garden St, Level 5, Eveleigh, 2015, NSW
Email: hoang.nguyen@data61.csiro.au

**Chen Cai**
Senior Researcher, Leader of Advanced Data Analytics in Transport group, DATA61| CSIRO
13 Garden St, Level 5, Eveleigh, 2015, NSW
Email: chen.cai@data61.csiro.au

Word count: 6200 words text + 6 figures x 250 words = 7700 words

A Paper Submitted for Presentation at the 97th Annual Meeting of the Transportation Research Board

**ABSTRACT:**

This paper introduces the framework of an innovative incident management platform with the main objective to provide decision support and situation awareness for transport management purposes on a real-time basis. The logic of the platform is to detect and then classify incidents into two types: recurrent and non-recurrent, based on their frequency and characteristics. Under this logic, recurrent incidents trigger the data-driven machine learning module which can predict and analyse the incident impact, in order to facilitate informed decisions for transport management operators. Non-recurrent incidents activate the simulation module which then evaluates quantitatively the performance of candidate response plans in parallel. The simulation output is used for choosing the most appropriate response plan for incident management. The current platform uses a data processing module to integrate complementary data sets, for the purpose of improving modelling outputs. Two real-world case studies are presented: 1) for recurrent incident management using data-driven model, 2) for non-recurrent incident management using traffic simulation with parallel scenario evaluation. The case studies demonstrate the viability of the proposed incident management framework which provides an integrated approach for real-time incident decision-support on large-scale networks.

*Keywords*: *Incident management, machine learning, data fusion, transport simulation, cloud-based data platform*

## 1. Introduction

An initiative of great value in the era of ‘Big Data’ is to effectively and consistently generate insights and actionable outcomes from multiple data sources. This would typically require a number of data handling processes, including data forwarding, parsing and integration, before data becoming usable by dedicated applications, say machine learning models. These data handling processes were conventionally deployed to enterprise data infrastructure, like the enterprise data warehouse (EDW). With the advent of cloud based infrastructure being epitomized by AWS and Azure, data handling, storage and computing are increasingly being integrated into a cloud data platform. Such a platform offers agility, scalability, and serviceability with substantial financial cost incentives, as compared to EDW solutions. Generic machine learning analytics are also being integrated with cloud data platforms, for the benefit of generating valuable insights on-demand. Examples of these are IBM Watson and Google Cloud Platform.

Of greater complexity from the generic cloud data platform are the domain platforms. The latter would require a considerable degree of knowledge on a particular domain, say urban transportation. Using the incident detection and classification as an example, which is further elaborated in the subsequent sections, a number of data sources would complement each other for improving the reliability and accuracy. Related data sources may include real-time traffic data, real-time incident monitoring data, crowd-sourced data, etc. Without a well-developed mathematical modeling framework that implements a number of fundamental principles, like the decision tree and support vector machines, it is unlikely the generic analytics would arrive at a deterministic and meaningful outcome.

In this paper, we present a part of the Advanced Data Analytics in Transport (ADAIT) platform (*1*). ADAIT is a cloud data platform hosted on AWS (Amazon Web Service) and dedicated to urban transport data analytics. In particular, we address the innovative approach of integrating data analytics with traffic simulation for incident decision-support. This approach exploits the distinction between recurrent and non-recurrent traffic incidents. With an adequate number of observations, information of a recurrent incident can be parsed into a number of prominent features and represented by a feature space. Data-driven models, i.e. machine learning models, could then be built to map the feature space into observed impact. The more observed repetitions of the recurrent incident, the better the data-driven model could explain the variance in the incident impact. On the other hand, for a non-recurrent incident, building a viable data-driven model may be prohibitive due to the lack of data. Instead, the detected incident information could be used to populate a simulation model. The latter can then run multiple scenarios in parallel, faster than real-time, to provide the expected impact of the incident. This paper discusses how the two modeling techniques could be integrated via the ADA IT platform to provide incident decision-support capabilities.

The rest of the paper is organised as follows: Section 2 presents the incident decision-support framework deployed in the ADAIT platform; Section 3 provides technical details for the key modules of the incident decision-support framework; Section 4 contains the case studies on handling recurrent incident and non-recurrent incident respectively; Section 5 summaries the findings of the paper.

## 2. Incident decision-support framework

The work presented in this paper focuses on incident detection and impact analysis on the traffic network and is a part of the ADAIT general framework, currently under development in collaboration with the traffic management center (TMC) in Sydney. FIGURE 1 presents the general flow chart of the incident detection framework, which comprises various types of interconnected modules, from the raw data processing and cleaning to the machine learning predictive module and traffic simulation core.

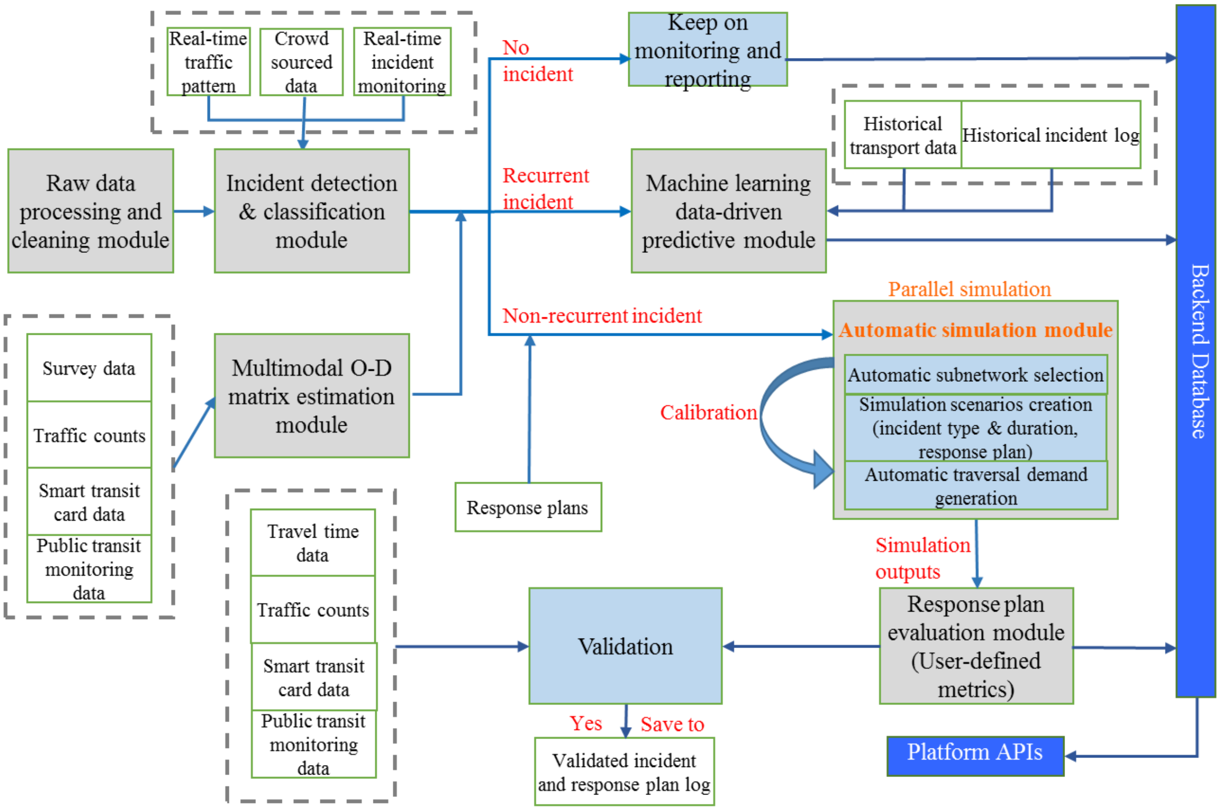



**FIGURE 1 Incident detection and impact analysis framework.**

The major characteristic of the ADAIT platform is that it offers a continuous situation awareness by real-time monitoring of the traffic condition, and triggers the data-driven pattern reconstruction or traffic simulation modules only when reported accidents may appear in the network. If no accident occurred, the platform will continue monitoring and reporting the traffic state in the network. All outputs of each module are stored in a dedicated back-end data base and the results are provided for consumption and further use through public APIs. Amongst the modules of the platform, we detail the following:

1. ***Incident detection and classification module*** (detailed in Section 3.3): which comprises: a) real-time data fusion from various sources for detecting recent incidents, b) incident ranking algorithms based on type and severity, and c) incident duration classification based on available characteristics of the reported accident. The main outcome of this module is to establish if the reported accident follows a recurrent pattern or not in order to efficiently trigger the simulation module for impact analysis and response plan evaluation.
2. ***Machine learning data-driven predictive module*** (see Section 3.4): which is based on historical observations of both transport and incident log data, and can predict the severity and duration of the detected incident.
3. ***Automatic simulation module*** (see Section 3.5): which is triggered only if the incident pattern is identified as non-recurrent. In this case, the incident feature information and a set of candidate response plans are passed to the simulation module. The module then automatically

selects the subnetwork where the incident takes place based on the incident feature information. It further generates a traversal demand matrix for the selected sub-network based on the path assignment obtained from previously running regular macroscopic static traffic assignment on the large-scale network. This will facilitate the application of mesoscopic or microscopic traffic scenarios at the subnetwork level for a more detailed simulation of the condition in the affected area while maintaining computation efficiency. Each response plan is simulated in parallel and the corresponding outcome will be evaluated according to user-defined performance metrics. Such parallel simulations can significantly reduce computation time, while providing quantitative insights on a mesoscopic or microscopic level. It is worth mentioning that this type of simulation module requires periodic calibration and validation of the simulation outputs by using existing data sources.

4. ***Multimodal O-D (Origin-Destination) matrix estimation module*** (see Section 3.2): it estimates the multimodal O-D demand of large-scale network based on historical transport data. The output of this module is a fundamental input for subsequent analysis and is critical to the simulation module's performance.

5. ***Response plan evaluation module:*** it chooses the best plan to mitigate the impact of the incident. The simulated outputs will be compared to the observed traffic data which will facilitate the transport operators to make informed decisions by various scenario testing and evaluation.

Combining these various modules with different characteristics and inputs/outputs, from data-driven incident detection/classification/prediction to automatic traffic simulation models, represents a unique and innovative method to evaluate the impact of incidents in a highly affected traffic network. The advantages of the proposed platform in the aforementioned perspectives are:

1. Integrating and fusing multiple data sources to improve the reliability and accuracy of results.
2. Applying machine learning algorithms to learn from previous incidents and make predictions.
3. Detecting and classifying incidents based on real-time data. The innovation is to only trigger the simulation model when the impact of a new incident could not be determined.
4. Automated simulation to provide quantitative evaluation of incident impact and response plan performance.

The logic and mechanism behind the platform are to provide a viable and practical solution to incident management. Such incident management system needs a detailed decomposition and theoretical analysis in terms of processing flow, traffic modeling, predictive analytics and response plan selection for mitigating the impact of traffic incidents on both normal and public transport modes. In the following, we will only focus on detailing the major modules for incident impact analysis and the way the information flow is propagated from one module to another with a major focus on the incident detection, duration prediction, and response plan generation.

# 3. Module details

In this section, we describe each module of the decision-support platform, as well as the implications and the afferent data sources needed for obtaining accurate insights.

## 3.1. Data processing module

The data sets used in the platform have a wide variety of formats and specifications, are often sparse and need constant cleaning and monitoring. Amongst them we cite:

a) *Survey data*: such as household travel survey data, census survey data etc., which are very time-consuming and become available every few years; they can provide insights on the O-D demand but may also be obsolete (*2*). On this platform, survey data is processed and provided to the multimodal O-D matrix estimation module as prior estimates of O-D matrix.

b) *Traffic counts*: are provided by the SCATS (Sydney Coordinated Adaptive Traffic System) system which plays a critical role in the multimodal O-D matrix estimation module, machine learning data-driven predictive module and validation. The integration of SCATS data into the platform is challenging due to a high complexity and duplication of streams, which needs supplementary processing, outlier detection, error elimination, etc.

c) *Smart transit card data*: provides tap-on and tap-off information of each user travelling in the city by public transport modes.

d) *Public transit monitoring data*: includes information such as fixed-route schedules, routes, and bus stop data. The GTFS (General Transit Feed Specification) provides such information on the Sydney network and is used for the validation of simulation results.

e) *Crowd-sourced data*: such as Twitter and Waze data are used on the platform as they provide textual information on incidents, and is processed directly by the incident detection and classification module.

f) *Real-time incident monitoring data*: provided by the transport management centre which reports incidents and actions taken on a real-time basis. This data is used by the Incident detection and classification module.

g) *Travel time data*: can be provided by many ITS applications; on the current platform the Google Travel Time is used for the validation of the simulation outputs.

## 3.2. Multimodal O-D matrix estimation module

The origin-destination demand matrix, which represents the number of trips from one urban centroid to another, is a fundamental element in many transport models. Having a reliable and accurate O-D matrix can significantly enhance the prediction quality (*3; 4*). In this paper, road traffic demand and public transport demand are estimated independently, and data fusion techniques are applied in both estimations. Multimodal O-D matrix estimation algorithms accounting for modal choice require additional information on behavioural analysis and choice modelling (*5; 6*) and is not adopted here due to computation complexity and data availability. Researchers have proposed various methods for data fusion in public transport (*7; 8*), even the traditional road traffic O-D estimation is an example of data fusion- the survey data is used as prior estimates of O-D matrix and is calibrated based on traffic counts (*9*). However, the integration of the multi-modal O-D estimation into the automated incident management platform is a novel approach which offers valuable input for the simulation module.

## 3.3. Incident detection and classification module

### *3.3.1 Incident detection*

Incident detection is considered an important component of many modern intelligent transport systems. The availability of data from multiple sources allows combining evidence from different systems in order to improve the detection performance. Multiple sources may provide complementary data, and multi-source data fusion can produce a better understanding of the observed situation by decreasing the uncertainty related to the individual (*7*). Traffic control operator can set a threshold to trigger and clear an incident alert using the alert score. The method to estimate the alert score is explained below in Equation (1), while the data used in this paper is explained in Section 3.1.

By analyzing the historical data sources along with confirmed incident logs, the important factor of the individual source is evaluated using attribute ranking algorithms (e.g. information gain, principal component analyses) (*10*).

Besides the reliability of individual source, the alert score is dependent on spatial and temporal aspects. For example, an accident happened on a major road during peak hour should have a higher score than a minor incident on a small road. Furthermore, a recent detection (e.g. within 10 minutes) should have more attention than past report (e.g. reported over 30 minutes). As a consequence, the spatial-temporal adjusted weight should be considered. An alert is generated if this score is higher than a threshold and cleared if it is below the threshold. The total alert score for an incident over the time is calculated as:

$$score(t) = \sum_{i=1}^{n} R_i W_i^j S(r) T(t) \tag{1}$$

Where $R_i$ is the reliability score for source $i^{th}$, $W_i^j$ is the weight for subtype of source $i$, $S(r)$ is the spatial adjustment function which depends on road type $r$ and $T(t)$ is the temporal adjustment which depends on time $t$.

*3.3.2 Incident classification*

When an incident is detected and confirmed by the system, a classification process is applied based on the incident description to divide it into different categories including accident, breakdown, delay, etc. Historical incident data is also used to train machine learning incident classification model to further classify incidents based on duration and severity. Duration is the time between the start and the end of an incident. Severity is a standard field in an incident log which is decided by a transport management operator to indicate the level of impact on traffic.

For incident duration classification, our system applies the method proposed by (*11*) to predict the duration of an incident that has just happened by using the available characteristics known at the onset, e.g. location, time, type of incident, lanes affected, operator in charge etc. Understanding the estimated incident duration is useful to traffic operators when choosing an appropriate response plan. Estimated incident duration is also one of the inputs for the traffic simulation module. Although several machine learning detection methods (*12; 13*) have been applied to detect an incident, few of them combined advanced machine learning, active learning and outlier detection techniques and achieves approximately 90% accuracy in predicting incident severity (*14*). This severity classification approach is integrated into our proposed system.

Finally, the detailed incident impact on the road network is then predicted by machine learning approaches (for recurrent incidents/congestions) or simulation modules (for unseen incidents).

### 3.4. Machine learning data-driven predictive module

When incidents occur in an urban traffic network, they are likely to affect the traffic flows of surrounding areas, especially to all the traffic leading to the congested roads. Some of the casual congestions follow the same patterns or sequences over the time. Therefore, it is useful to discover frequent patterns (if any) of congestion propagations by reviewing historical data in the traffic networks.

This section reviews an algorithm that finds congestion propagation pattern by looking at the relationships of congestions from the earliest data record through the latest one (*15*). The main insight is that congestion $C_1$ is a parent of congestion $C_2$ if $C_1$ occurred before $C_2$ in time and they are spatially connected.

The subfigure on the left illustrates top 3 congested segments in three consecutive snapshots, and the one on the right shows two causal congestion trees obtained from them. A snapshot is the state of the traffic condition from all segments in the network at a given time. Each tree is presented by a list of connected segments. A frequent tree represents expected congestion propagation pattern when an incident happens on a root segment (e.g. B→A segment). A root of a congestion tree is defined as a segment where traffic from other segments are flowing into it and causing congestion.

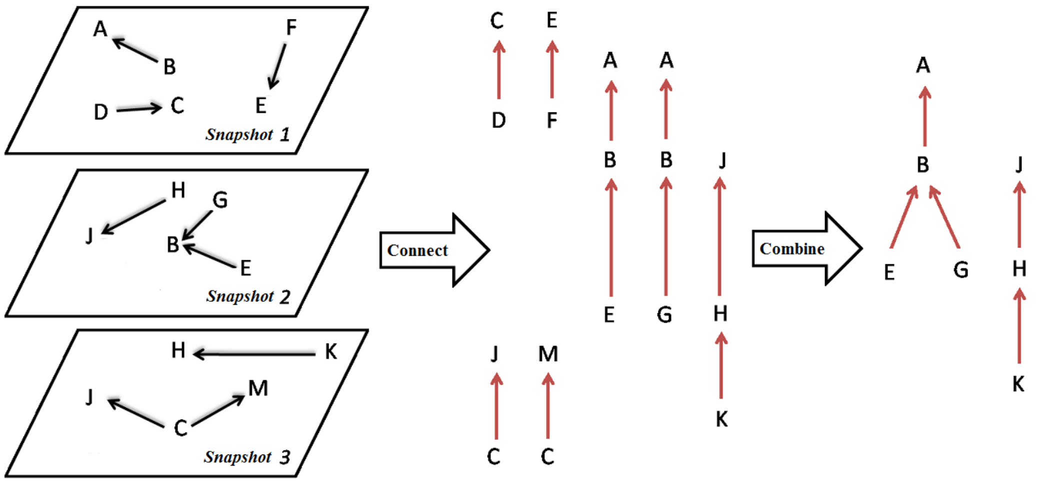



**FIGURE 2 An example for demonstrating the process of building a forest of two Causal congestion trees.**

A Dynamic Bayesian Network (DBN) approach is used to model the spatial-temporal characteristics of a recurrent congestion propagation. Using this method, the probability of a recurrent incident's impact can be estimated. DBN is a popular approach for modelling spatial-temporal data (HN 11). A DBN is usually referred as a 2-Time slice Bayesian Network (2TBN) because at any given time T, the value of a variable is computed from the internal regressors and the immediate prior value (time T-1) (*16*). Therefore, DBN is reasonably close to the real-world phenomenon of traffic congestion where the status of a segment at a specific time can be determined by its previous condition and previous conditions of connecting segments.

To build the DBN traffic network, the road segments are presented by a set of $N_h$ random variables, $O_t^{(i)} \in [0,1]$, where 1 represents a congested segment at time t. Snapshot t is a storage of the traffic condition from all segments in the network at time t.

In a DBN, the transition (denoted as $B_{\rightarrow}$) and observation model $P(Z_t|Z_{t-1})$ is then defined as a product of the conditional probability distribution (CPD) in the 2TBN:

$$P(Z_t|Z_{t-1}) = \prod_{i=1}^{N} P(Z_t^{(i)}|Pa(Z_t^{(i)})) \tag{2}$$

Where $Z_t^{(i)}$ is the $i^{th}$ node in snapshot $t$ and $Pa\left(Z_t^{(i)}\right)$ are the parents of $Z_t^{(i)}$. The unconditional initial state distribution $P(Z_1^{1:N})$ is presented by a standard Bayesian Network, namely $B_1$. Together, $B_1$ and $B_{\rightarrow}$ define the DBN.

Suppose we have a simple traffic network which comprises of three segments: EB and GB are connected to BA. As EB and GB both lead to BA, when BA is congested, it becomes the potential cause for congestions at EB and GB in the next time frame. The corresponding DBN network is presented in FIGURE 3.

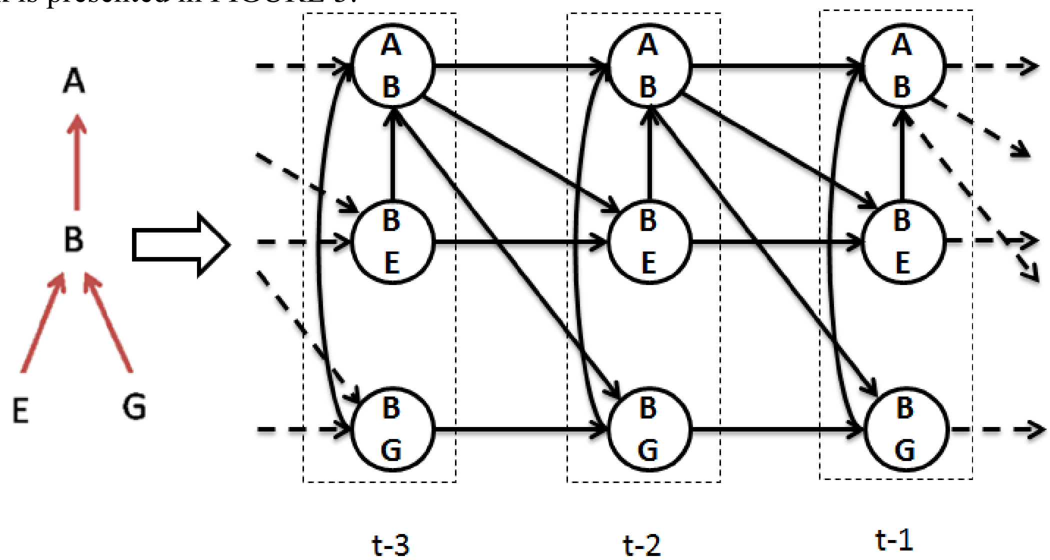



**FIGURE 3 Modelling congestion propagation of a 3-segment traffic network by DBN.**

When the propagation pattern is generated, the joint distribution for a known-structure tree which includes *T* consecutive snapshots (slices) can be obtained by "unrolling" the network until we have *T* slices, and then multiplying together all of the conditional probability distribution.

$$P\left(Z_{1:T}^{(1:N)}\right) = \prod_{i=1}^{N} P_{B_1}(Z_1^{(i)}|Pa(Z_t^{(i)}) \times \prod_{t=2}^{T}\prod_{i=1}^{N} P_{B_\rightarrow}(Z_1^{(i)}|Pa(Z_t^{(i)}) \quad (3)$$

In case the detected incident belongs to the root of congestion propagation tree and the probability to form a propagation pattern is higher than a predefined threshold (e.g. 60%), the traffic controller may decide to rely on this impact pattern to control the traffic rather than executing simulation which is more time-consuming. In case there is no propagation that exceeded the predefined threshold, the simulation technique is applied to test the impact of an incident on the road network.

### 3.5. Simulation module

In the real world, the location, type, and severity of an incident may vary significantly and hence a number of incidents may not be accurately and reliably predicted by the machine learning data driven predictive module. In this case, the simulation module needs to be activated to evaluate the impact of the incident quantitatively. However, simulation of large-scale networks in a reasonable time can hardly be viable due to the computational complexity. Some real-time traffic simulation models have opted for simulation at a corridor/motorway level (*17-19*) but few offer automatic sub-network selection and real-time incident simulation based on duration prediction

Therefore, we have proposed an automated parallel simulation module to address these issues. In this module, a large-scale traffic simulation model is constructed for the city of Sydney, Australia, and a macroscopic assignment process is implemented every day based on the new incoming data received from the data processing module. Macroscopic assignment models have been well studied by many researchers and many existing models are available, such as the stochastic user equilibrium assignment (*20*), static user equilibrium assignment (*21; 22*), system optimal assignment (*23*) and many variations and extensions (*24; 25*). The platform can apply these models depending on these models' applicability, which shows the platform's flexibility and extensibility. In the case study, the algorithm by Florian is used (*25*).

To achieve a reasonable computation time, the large-scale network is sub-divided into small subnetworks. For example, the Bureau of Transport Statistics in New South Wales, Australia, applies a geometrical zoning configuration which is open to the public (*26*). In addition, some research papers also provide methods for automatic zoning when zoning information is not available (*27*). Based on the incident information (location, severity, lanes affected), the module automatically selects the afferent subnetwork, where further analysis and prediction are applied. Then, the traversal demand matrix is generated for the selected sub-network. Here, the traversal demand matrix represents the travelers that travel through, in and out of the subnetwork. To further reduce the computation time, each response plan is simulated simultaneously by the parallel simulation module. This is because transport operators can have multiple candidate response plans, and the simulation module can facilitate them to choose the most suitable plan. On this platform, the mesoscopic simulation has been adopted because it is computationally more efficient than microscopic simulation, less data-demanding while capturing the essentials of the traffic dynamics (*28-30*). In cases where microscopic-level simulation is necessary, models accounting for car-following behavior (*31*), lane-changing behavior (*32; 33*) and gap acceptance (*34*) may also be used in the module. Note that all the aforementioned simulation processes are automated from the beginning, which is also critical to reducing the computation time to satisfy the strict requirement on response time for incident management purposes.

The reliability of this decision-making process relies on a simulation model that represents the system's behavior closely enough (*35; 36*), and hence the simulated outputs are validated with several data sources: travel time data (e.g. probe vehicle data, taxi data or Google travel time data), traffic counts (e.g. loop detector data) and public transport data (e.g. smart card data and public transport monitoring data). If the results do not closely approximate the observations, further calibration is conducted to fine tune the parameters such as sub-network traversal demand, mesoscopic reaction time and so forth until the result quality is satisfactory.

### 3.6. Response plan evaluation module

As each city or traffic management centre has its own unique characteristics and preferences, it is important that the best response is chosen based on bespoke metrics, so that the ramification of incident is mitigated in a user-defined way. For example, the average travel time per kilometre can be used to evaluate performance:

$$TT_{network} = \frac{\sum_{i=1}^{N_{network}} TT_i}{N_{network}} \tag{4}$$

Where, $TT_i$ is the average travel time per kilometre of vehicle $i$, $N_{network}$ is the number of vehicles in the network. When travel time reliability or other factors are considered important, they may also be used to choose the most suitable response plan. Such a user-defined metric embodies the platform's flexibility and customizability.

## 4. Case study

In this section, we present a case study of the Sydney large-scale network consisting of more than

70000 links and 2000 centroids. This case study focuses mainly on the incident detection and classification, recurrent incident impact prediction and non-recurrent incident simulation. The data sources used on this platform have been explained in Section 3.1.

### 4.1. Recurrent incident: congestion propagation and impact prediction

When an incident is detected, the incident detection and classification module will also analyze and determine whether the incident is recurrent or not. In this subsection, an example of recurrent incident is presented, to demonstrate that the machine learning data-driven predictive module can predict the congestion propagation, incident impact for transport operators to make informative decisions.

At 8:30 am on a weekday, an accident was detected at George St near Campbell St in the Sydney CBD. Given that the accident first happened at the root segment BA, using congestion discovery algorithm and DBN (Section 3.4), a 5-segment congestion propagation pattern was detected with a joint distribution probability estimated at 74%. This probability is higher than the predefined threshold of 60% hence this predicted impact can be used by the operator to manage the incident.

The detected congestion pattern was then validated using the real-time traffic data collected from SCATS. The congestion was initially detected at segment BA. Five minutes later, both segments DB and CB also became congested. Until 8:45 am which is 15 minutes after the incident happened, similar congestion propagation patterns were detected on all five segments. The case study shows the efficiency and capability of the machine learning module in decision support and impact prediction for recurrent incidents.

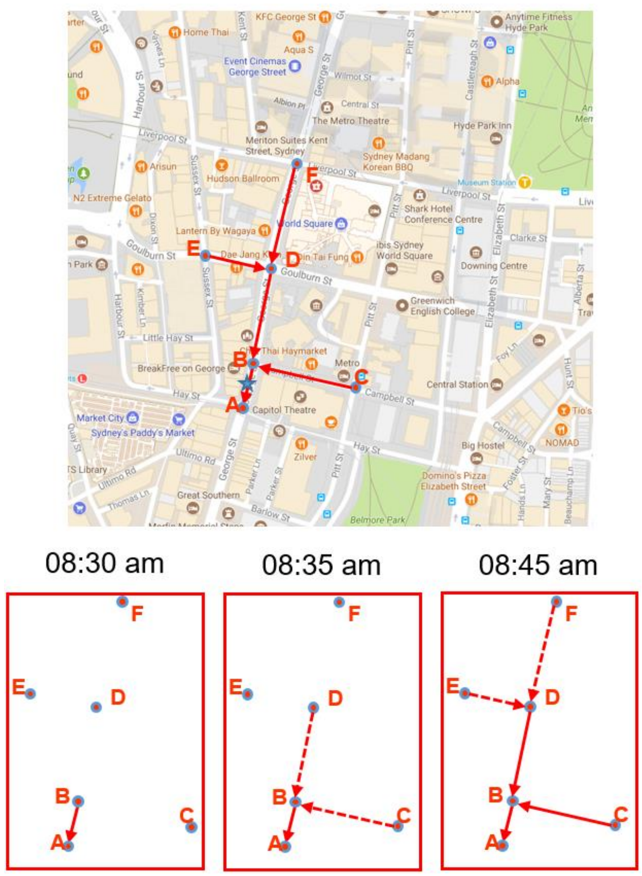



**FIGURE 4. The frequent congestion propagation pattern in the Sydney CBD.**

### 4.2. Non-recurrent incident: parallel simulation and performance evaluation

Although the machine learning data-driven predictive module shows its efficiency in predicting incident impact, a number of incidents may not be predicted correctly when little historical information is available. The simulation module needs to be triggered to assist transport operators when such non-recurrent incidents are detected. As previously explained, simulation of the whole Sydney network on a Mesoscopic or Microscopic level is extremely time-consuming, which cannot satisfy the computational time requirements for transport management purposes. Therefore, the Sydney network has been sub-divided into many sub-networks beforehand using the Statistical Area definition in (*26*), so that the simulation module can select a sub-network promptly for simulation. The transport simulation model of the Sydney network is implemented in AIMSUN, which is regularly calibrated using periodically aggregated SCATS traffic counts and smart card data. The AIMSUN network simulation model uses the multimodal O-D matrix previously estimated (see Section 3.2) and runs a macroscopic multimodal traffic assignment for the Sydney network. The output will be saved for later use in generating a traversal demand for the chosen sub-network where the incident happens.

The non-recurrent incident presented here is reported by the incident detection and classification module with the following information:

a. Location (including x and y coordinates): Pyrmont bridge road, Pyrmont.
b. Estimated duration: 30 minutes.
c. Severity: major accident affecting all lanes in both directions.
d. Start time: 07:15 a.m.
e. Incident pattern: non-recurrent.

The non-recurrent incident pattern triggers the simulation module. To demonstrate the network state and performance before and after the incident, a 2-hour simulation period is chosen, from 7:00 a.m. to 9:00 a.m. Using the incident location, a subnetwork has been automatically selected from the list of available subnetworks in the city, which is identified as -Pyrmont. Pyrmont is a suburb adjacent to Sydney CBD and the majority of its area is zoned for commercial purposes. See FIGURE 5:

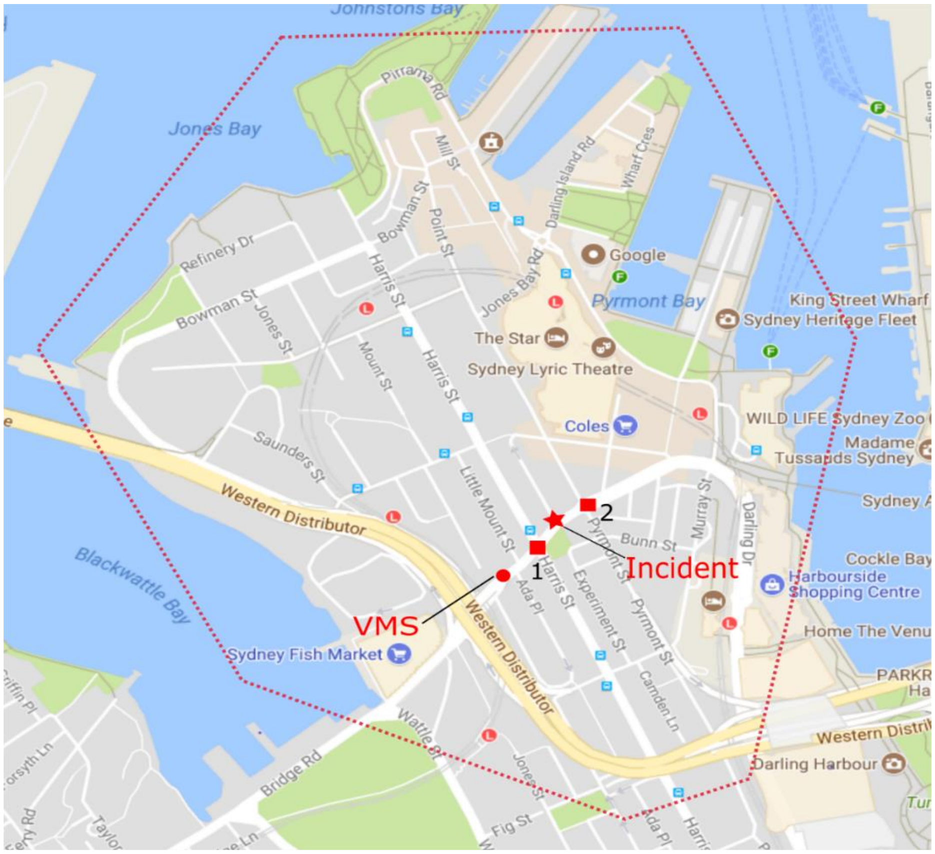



**FIGURE 5 The selected sub-network (Pyrmont, New South Wales).**

After automatically selecting the sub-network area in which the incident has occurred, the traversal demand matrix for Pyrmont is generated and calibrated for the morning rush hour (7:00 a.m. to 9:00 a.m.). The model is also validated by comparing the average travel time obtained from simulation (STT) on each road section with the average travel time from Google (GTT) or the average travel time obtained from the SCATS data in Pyrmont (SCATSTT). FIGURE 6a) presents an example of comparison between the average STT and GTT on the road section 2839_2840 from Pyrmont, on a Wednesday morning from 7 to 9 AM. The plot of travel time every 15 minutes indicates that the simulation provides good results of the TT on this section as it falls between the $5^{th}$ and $95^{th}$ percentile of the GTT. This finding is validated once more on a different section (5_2839), where the STT is compared to GTT SCATSTT which is available for computation (FIGURE 6b).

Based on the received incident information, various incident response plans (RPs) are generated for evaluation:

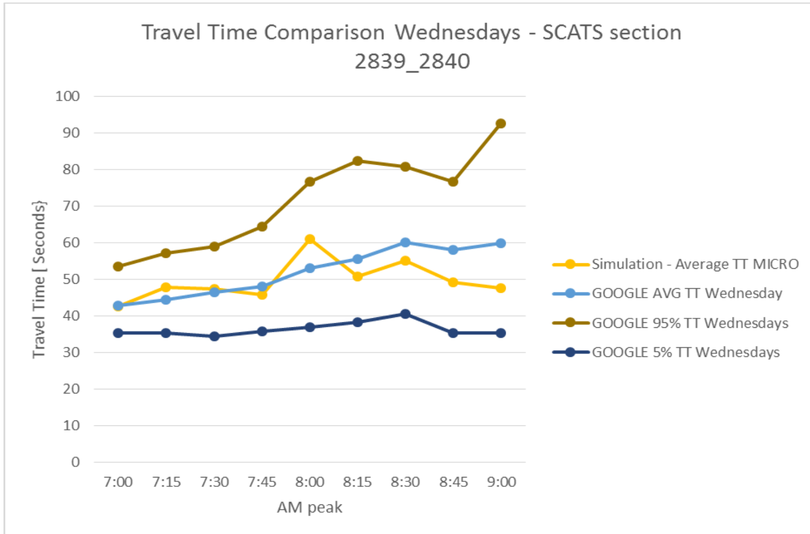
1) RP1- Do nothing.
2) RP2- Redirect all traffic in intersections 1 and 2 (marked as red rectangular in FIGURE 5 towards adjacent intersections.
3) RP3: Combined actions: Activate the VMS to redirect all off-ramp flow from the bridge towards Little Mount St. (see FIGURE 5), and redirect all traffic in intersection 1 towards surrounding intersections.
4) RP4: Activate the VMS to redirect all off-ramp flow from the bridge towards Little Mount St.

RP1 is intended to keep monitoring the network but take no action, in order to evaluate the true impact of the accident if traffic operators would not react to the accidents. RP2 redirects all traffic in intersections 1 and 2 towards adjacent intersections in order to prevent vehicles from queuing and eventually blocking the intersection. RP3 has the role to activate the VMS (Variable message sign), which will inform drivers to make a left turn before reaching intersection 1. Also, traffic will be redirected at intersection 1 to prevent queuing. RP3 aims to let the major traffic from the bridge (Western distributor) bypass intersection 1. RP4 simplifies RP3 by keeping only the VMS activation action.

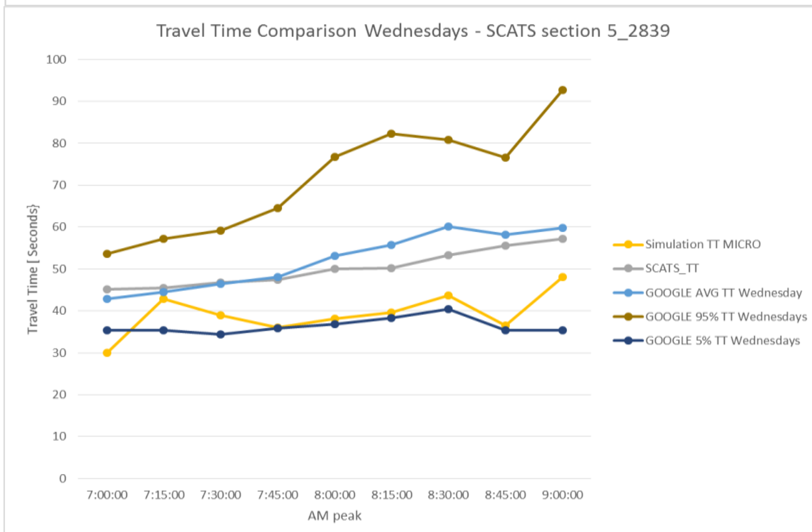
These response plans are then simulated in parallel on a microscopic level by the microsimulation engine in AIMSUN. FIGURE 6c) presents the average travel time per kilometre (including bus and private vehicles) obtained after applying each of the 4 response plans. Plan 1 is the baseline and demonstrates that the average travel time reaches a high point at 8:00 a.m., when the incident has already ended. This is due to the accumulation of queue and the increase in traffic and public transport demand. Plan 4 shows a very marginal improvement over plan 1, the average travel time over the 2 hours is also quite close to plan 1. Although plan 2 mitigates the congestion during 7:45 a.m. to 8:15 a.m., the travel time increases after 8:15 a.m. gradually, making the eventual average travel time over the 2 hours very similar to response plan 1. Overall, plan 3 performs the best, it smooths the travel time after the incident happens, while having a 7% reduction in average travel time over the 2 hours simulation period.

The finding indicates that the best response plan for mitigating congestion produced by a non-recurrent incident is actually a combination of various actions which complement each other and help to reduce the incident clearance time. Therefore a possible extension of this work is to automatically recommend the best combination of response plans to apply for efficiently easing congestion.

)




)




)

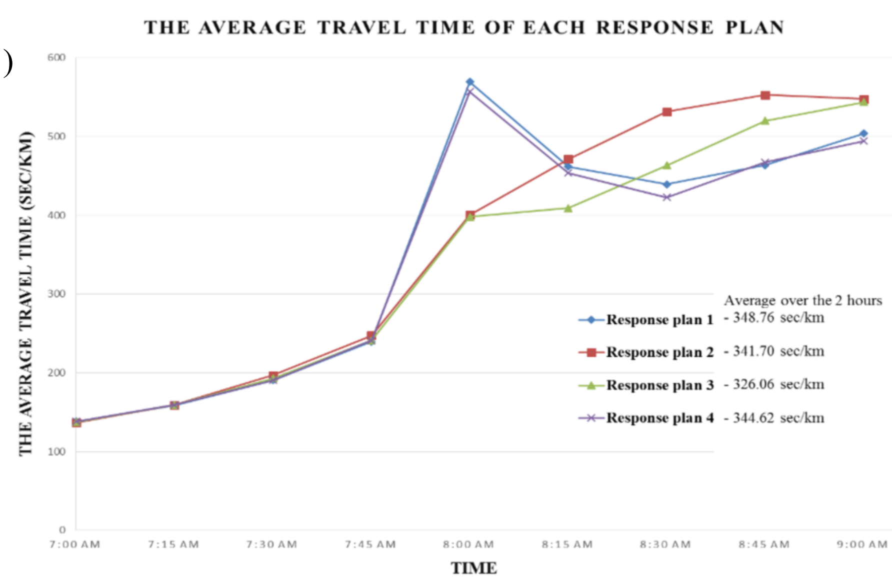



**FIGURE 6. Average travel-time comparison between: a) simulation and Google Travel Time, b) simulation, Google TT and SCATS c) each response plan (RP1, RP2, RP3, RP4).**

## 5. Conclusion

In this paper, we introduced the general framework of the ADAIT platform and explained the main function of each module. The platform can detect and then classify incidents into recurrent and non-recurrent pattern, the former one triggers the machine learning data-driven predictive module which predicts the incident duration and impact, so transport management operators can decide if the simulation module needs to be activated. Non-recurrent incident is directly passed to the simulation module, the performance of candidate response plans is evaluated quantitatively, and then operators can opt for the best plan to mitigate the negative impact of an incident. Case studies demonstrate that the impact of recurrent incident, such as congestion propagation, can be predicted by the machine learning module, and the simulation module can help choose the best response plan to mitigate the negative incident impact. In short, data-driven incident detection/classification, machine learning analytics for incident prediction and automatic traffic simulation models are integrated into the cloud-based platform, which represents a unique and innovative method to evaluate the impact of incidents in real-time for large-scale networks.

There are various possible opportunities to further extend the platform: The response plans can be generated automatically by advanced machine learning techniques based on the information of a detected incident (such as location, duration, severity etc.), and hence manual input of response plan is no longer required from operators. Also, the platform's modularity allows integration of advanced transport algorithms in each module, which can enhance the platform's applicability.

## 6. Acknowledgement

The work presented in this paper is partially funded by the New South Wales Premiere's Innovation Initiative. The authors of this work are grateful for the work and support of the Traffic Management Centre from Transport for New South Wales Australia. Data61 is funded by Australian Federal Government through Commonwealth Scientific and Industrial Research Organisation.

**REFERENCES**

1. ADAIT(Advanced Data Analytics in Transportation). Data61, CSIRO. www.adait.io.
2. Cascetta, E. *Transportation systems analysis: models and applications*. Springer Science & Business Media, 2009.
3. Bierlaire, M., and P. L. Toint. Meuse: An origin-destination matrix estimator that exploits structure. *Transportation Research Part B: Methodological,* Vol. 29, No. 1, 1995, pp. 47-60.
4. Bell, M. G., and Y. Iida. Transportation network analysis. 1997.
5. Pattanamekar, P., and D. Park. Estimating multimodal OD matrix from mode specific OD matrices.In *Transportation Research Board 88th Annual Meeting*, 2009.
6. Zhang, L., C. Xiong, and K. Berger. Multimodal inter-regional origin-destination demand estimation: A review of methodologies and their applicability to national-level travel analysis in the US WCTR. *Lisbon, Portugal, July*, 2010.
7. El Faouzi, N.-E., H. Leung, and A. Kurian. Data fusion in intelligent transportation systems: Progress and challenges–A survey. *Information Fusion,* Vol. 12, No. 1, 2011, pp. 4-10.
8. Kusakabe, T., and Y. Asakura. Behavioural data mining of transit smart card data: A data fusion approach. *Transportation Research Part C: Emerging Technologies,* Vol. 46, 2014, pp. 179-191.
9. Hazelton, M. L. Some comments on origin–destination matrix estimation. *Transportation Research Part A: Policy and Practice,* Vol. 37, No. 10, 2003, pp. 811-822.
10. Guyon, I., and A. Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research,* Vol. 3, No. Mar, 2003, pp. 1157-1182.

11. Taib, R., D. Yee, F. Chen, and W. Liu. Use of Machine Learning for Improved Incident Management Outcomes.In *the 13th ITS Asia Pacific Forum 2014, Auckland, New-Zealand, 2014*, 2014. p. 10.
12. Chong, M., A. Abraham, and M. Paprzycki. Traffic accident analysis using machine learning paradigms. *Informatica,* Vol. 29, No. 1, 2005.
13. Sohn, S. Y., and H. Shin. Pattern recognition for road traffic accident severity in Korea. *Ergonomics,* Vol. 44, No. 1, 2001, pp. 107-117.
14. Nguyen, H., C. Cai, and F. Chen. Automatic classification of traffic incident's severity using machine learning approaches.In *IET Intelligent Transport Systems*, Institution of Engineering and Technology, 2017.
15. Nguyen, H., W. Liu, and F. Chen. Discovering congestion propagation patterns in spatio-temporal traffic data. *IEEE Transactions on Big Data,* Vol. 3, No. 2, 2017, pp. 169-180.
16. Ghahramani, Z. Learning dynamic Bayesian networks.In *Adaptive processing of sequences and data structures*, Springer, 1998. pp. 168-197.
17. Transport Simulation Systems, T. Integrated Corridor Management in San Diego, California. https://www.aimsun.com/integrated-corridor-management-project-in-san-diego/, 2012.
18. Dia, H., and N. Cottman. Evaluation of arterial incident management impacts using traffic simulation.In *IEE Proceedings-Intelligent Transport Systems, No. 153*, IET, 2006. pp. 242-252.
19. Hidas, P. Modelling vehicle interactions in microscopic simulation of merging and weaving. *Transportation Research Part C: Emerging Technologies,* Vol. 13, No. 1, 2005, pp. 37-62.
20. Zhang, K., H. S. Mahmassani, and C.-C. Lu. Probit-based time-dependent stochastic user equilibrium traffic assignment model. *Transportation Research Record*, No. 2085, 2008, pp. 86-94.
21. Beckmann, M., C. McGuire, and C. B. Winsten. Studies in the Economics of Transportation.In, 1956.
22. Wen, T., C. Cai, L. Gardner, V. Dixit, S. T. Waller, and F. Chen. A Strategic User Equilibrium for Independently Distributed Origin-Destination Demands.In *Transportation Research Board 95th Annual Meeting*, 2016.
23. Chow, A. H. F. Properties of system optimal traffic assignment with departure time choice and its solution method. *Transportation Research Part B: Methodological,* Vol. 43, No. 3, 2009, pp. 325-344.
24. Patriksson, M. *The traffic assignment problem: models and methods*. Courier Dover Publications, 2015.
25. Florian, M., and D. Hearn. Network equilibrium models and algorithms. *Handbooks in Operations Research and Management Science,* Vol. 8, 1995, pp. 485-550.
26. Australian Bureau of Statistics, A. Main Structure and Greater Capital City Statistical Areas, July 2011 In *Australian Statistical Geography Standard (ASGS): Volume 1*, Institution of Engineering and Technology, 2017.
27. Menon, A. K., C. Cai, W. Wang, T. Wen, and F. Chen. Fine-grained OD estimation with automated zoning and sparsity regularisation. *Transportation Research Part B: Methodological,* Vol. 80, 2015, pp. 150-172.
28. Mahut, M., M. Florian, N. Tremblay, M. Campbell, D. Patman, and Z. McDaniel. Calibration and application of a simulation-based dynamic traffic assignment model. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1876, 2004, pp. 101-111.
29. Burghout, W., H. Koutsopoulos, and I. Andreasson. Hybrid mesoscopic-microscopic traffic

simulation. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1934, 2005, pp. 218-255.

30. Ben-Akiva, M., M. Bierlaire, H. N. Koutsopoulos, and R. Mishalani. Real time simulation of traffic demand-supply interactions within DynaMIT.In *Transportation and network analysis: current trends*, Springer, 2002. pp. 19-36.
31. Gipps, P. Multsim: a model for simulating vehicular traffic on multi-lane arterial roads. *Mathematics and Computers in Simulation,* Vol. 28, No. 4, 1986, pp. 291-295.
32. Ahmed, K., M. Ben-Akiva, H. Koutsopoulos, and R. Mishalani. Models of freeway lane changing and gap acceptance behavior. *Transportation and traffic theory,* Vol. 13, 1996, pp. 501-515.
33. Kesting, A., M. Treiber, and D. Helbing. General lane-changing model MOBIL for car-following models. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1999, 2007, pp. 86-94.
34. Farah, H., S. Bekhor, A. Polus, and T. Toledo. A passing gap acceptance model for two-lane rural highways. *Transportmetrica,* Vol. 5, No. 3, 2009, pp. 159-172.
35. Dowling, R., A. Skabardonis, J. Halkias, G. McHale, and G. Zammit. Guidelines for calibration of microsimulation models: framework and applications. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1876, 2004, pp. 1-9.
36. Barcelo, J., and J. Casas. Methodological notes on the calibration and validation of microscopic traffic simulation models.In *Proc. 83rd annual meeting of transportation research board annual conference, Washington DC*, 2004.