Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

International Symposium of Transport Simulation (ISTS'18) and the International Workshop on Traffic Data Collection and its Standardization (IWTDCS'18)

Traffic Signal Optimisation in Disrupted Networks with Re-routing

Dana Abudayyeh[a*], Dong Ngoduy[a], Alan Nicholson[a]

[a]*Department of Civil and Natural Resources Engineering, University of Canterbury, Christchurch 8140, New Zealand*

Abstract

This paper describes a modified bi-level framework to simulate disruptive events in urban road networks with different levels of disruption severity and duration. This framework combines the Cross Entropy method to optimise traffic signals and the quasi-dynamic user equilibrium assignment model embedded in SATURN software package. This enables simulation of short-term closures with less computational effort and running time than fully dynamic models. We have applied this model to the Cambridge (UK) network and demonstrated how the degradation at one node affects the optimal signal settings at that node and nearby nodes. The computational results show that for different disruption severities, as traffic starts to divert to other routes, the optimal traffic signal settings changes, to minimise the travel time.

Keywords: Disrupted road network; network reliability; Cross Entropy method; traffic signal optimisation; quasi –dynamic; resilience

1. Introduction

To ensure operational continuity of urban road networks, a transportation system's resilience has become an important issue. Over the last two decades, there has been extensive discussion about the need for robust networks to minimise the economic and social impacts of disruptions. Detailed reviews of the literature related to degraded networks have been conducted, e.g. Berdica (2002) and Mattsson and Jenelius (2015). Koorey et al. (2015) explored the scope for dynamic traffic signal control to reduce the impact of disruptions associated with non-recurrent congestion (e.g. traffic incidents). It has been suggested that reducing these will have a great effect on network reliability as half of the congestion is caused by non-recurring events (Pearce, 2000, Schrank et al., 2009).

* Corresponding author. Tel.: +64-2040749702.
E-mail address: dana.abudayyeh@pg.canterbury.ac.nz

2352-1465

Several studies of infrastructure resilience have proposed a disruption profile to capture the phases of any significant disruption before, during and after the disruption. For example, Asbjornslett (1999) proposed three phases, i.e. stable situation before a disruptive event, disruption time, and a new stable situation after the disruption time has passed (Fig. 1a). The new stable situation may be better or worse than the one before the disruption. Sheffi (2005) identified five typical phases of the disruption profile (Fig. 1b): the preparation phase, the disruptive event, the first response, the recovery preparation, and the recovery. Both authors indicated that the severity of a disruptive event dictates the initial network performance reduction and the recovery time. Additionally, Bruneau et al. (2003) have proposed a resilience triangle (Fig.1c), suggesting that the smaller the area of the triangle the greater the resilience. More recently, Taylor (2017) presented another representation to reflect the dynamic performance of an infrastructure system (Fig.1d). This distinguishes between frequent minor variations in performance and infrequent major disruptions.

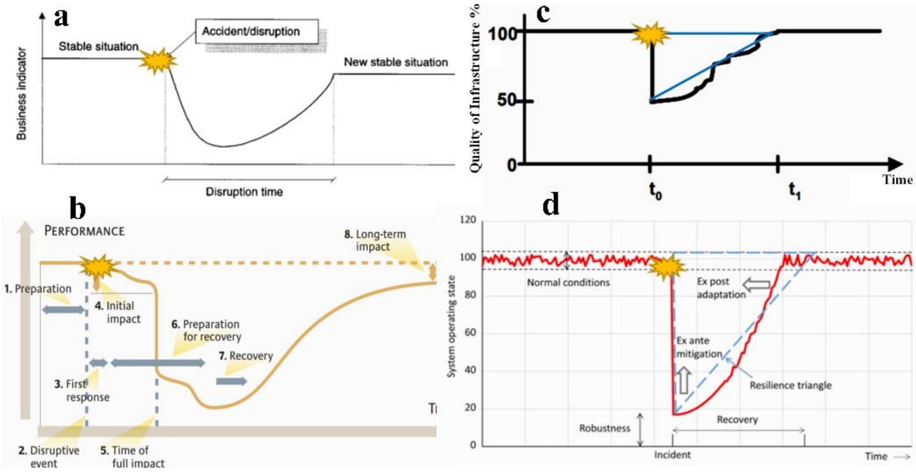

Fig. 1.(a) Regaining stability after a disruption (Asbjornslett, 1999); (b) Sheffi's disruption profile (Sheffi, 2005); (c) The concept of resilience triangle (Bruneau et al., 2003); (d) The resilient triangle in traffic (Taylor, 2017)

We are seeking to improve the resilience by reducing the size of the resilience triangle by reducing its height and/or base. There are various options to achieve this, including constructing or improving parallel routes between given pairs of nodes. An alternative option is to use traffic signal control. The aim of this study is to reduce the impact of a disruptive event (i.e. infrequent variations) using traffic signal control, as previously investigated by Koorey et al (2015).

Traffic signal control can be used to assist drivers to avoid blockages and to use other routes to minimise delays. Various optimisation algorithms have been implemented to find the optimal set of signal timings, taking into account the impact of re-routing. One of these optimisation methods is the Cross Entropy (CE) method proposed by Rubinstein (1997). Maher (2008) introduced the CE algorithm to optimise the signal settings on a six-arm signalised roundabout. Ngoduy and Maher (2011) and Maher et al. (2013) further explored the CE method to optimise traffic signals in urban networks. The results of applying the CE method showed encouraging advantages for computational efficiency and convergence, with its more formal mathematical and statistical basis making it simple to apply (Maher, 2008), as was also found by Ngoduy and Maher (2012) and Zhong et al. (2016), who used the CE method to calibrate microscopic traffic models. Maher (2008), Ngoduy and Maher (2011), and Maher et al. (2013) used a bi-level framework approach, as did Kaviani et al. (2017), who sought the optimal locations of roadside guidance devices across a regional road network for improving total travel time within a network during long-term closures such as natural disasters.

This paper describes a bi-level optimisation framework to minimise the impact of disruption (i.e. to minimise the travel time in a network) by changing the signal settings (i.e. the green times and offsets), to facilitate re-routing around blockages of various severities and durations.

2. Research method and implementation

To understand the impact of disruptions on traffic network performance under optimum signal control, a bi-level optimisation problem was formulated. The approach, which was introduced by Ngoduy and Maher (2011), was adopted and extended to account for urban network degradations. The process for optimising the signal settings involves iterating between the CE algorithm and SATURN. The CE algorithm searches for the combination of signal settings which minimises the total travel time, calling SATURN to estimate the flows and travel times for specified combinations of signal settings, considering re-routing. The iterative process continues until satisfactory convergence is achieved (Fig. 2).

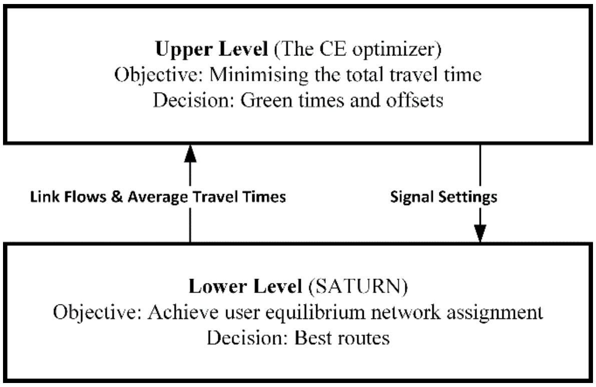

Fig. 2 Bi-level optimisation framework

2.1. The bi-level framework formulation

The upper level optimisation problem represents planners trying to minimise the average travel time immediately after the disruptive event, when equilibrium has not yet been reached among the road users. The upper level of the problem is formulated as:

$$Min\, PI(\boldsymbol{X}, \boldsymbol{q_{UE}}(\boldsymbol{X})) = \sum_{a=1}^{L} q_a t_a(\boldsymbol{X}, \boldsymbol{q_{UE}}(\boldsymbol{X})); \quad subject\, to: \boldsymbol{X}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{C}) \in \Omega \tag{1}$$

where $PI(X, q_{UE}(X))$ is the performance index function (i.e. the total travel time in the network) which depends on the vector of link equilibrium flows $\boldsymbol{q_{UE}}$ and the vector of signal timings $\boldsymbol{X}$ consisting of the vector of offsets $\boldsymbol{\beta}$, the vector of green times $\boldsymbol{\theta}$ and the cycle length $\boldsymbol{C}$; L is the number of links; q_a is the flow on link a; t_a is the average travel time for the link flow; Ω denotes the feasible space of X defined as:

$$C_{\min} \le C \le C_{\max}; \quad 0 \le \beta_n \le C-1; \quad \theta_{n,s}^{\min} \le \theta_{n,s} \le \theta_{n,s}^{\max}; \quad C = \sum_{s=1}^{S_n} \theta_{n,s} + \sum_{s=1}^{S_n} I_{n,s} \tag{2}$$

where $C_{\min}$ and $C_{\max}$ are the lower and upper bound of the cycle length, respectively; β_n is the offset at node n; $\theta_{n,s}$ is the green time at node n for stage s; $\theta_{n,s}^{min}$ and $\theta_{n,s}^{max}$ are the lower and upper bound of the green time at node n for stage s; S_n is the number of stages at node n; $I_{n,s}$ is the inter-green time at node n for stage s. We consider the signal settings to be discrete integer values.

The lower level represents users following the user equilibrium principle under the given network condition. This can be formulated as:

$$\boldsymbol{t}(\boldsymbol{X}, \boldsymbol{q_{UE}}).(\boldsymbol{q} - \boldsymbol{q_{UE}}) \ge 0 \qquad \forall q \in \Theta \tag{3}$$

where $\boldsymbol{q}$ is the vector of link flows and $\boldsymbol{q_{UE}}$ is the vector of equilibrium link flows. In equation (3), $\boldsymbol{t}(\boldsymbol{X}, \boldsymbol{q_{UE}})$ denotes the vector of link travel times, which is dependent on the vector of signal timings and the equilibrium link flows. Θ denotes the feasible space of the link flow vector and is explicitly defined as:

$$\sum_{p\in P} f_{ijp} = OD_{ij} \quad \forall i\in O, j\in D \quad ; \quad f_{ijp} \geq 0 \quad \forall i\in O, j\in D, p\in P$$
$$q_a = \sum_{i\in O}\sum_{j\in D}\sum_{p\in P} f_{ijp}\delta_{aijp} \quad \forall a\in L \quad ; \quad q_a \leq q_a^0 \quad \forall a\in L \tag{4}$$

where q_a^0 is the link capacity; O and D are the set of origins and destinations; P is the set of possible paths; i, j are the origin index and destination index; p is the path index; f_{ijp} is the path flow between origin i and destination j using path p; δ_{aijp} is an indicator variable which equals one if the link a is on path p between i and j, and zero otherwise.

2.2. SATURN mesoscopic simulator

In principle, microscopic packages are very suited for dealing with short-term traffic closures, for the reason that these models are able to calculate the optimal paths periodically and reassign vehicles to new optimal paths, to take account of route changes after a trip has begun, to avoid a blockage to minimise delay. In the absence of disruption or congestion no re-assignment will occur. This regular updating is more appropriate for studying short-term capacity reductions. However, a major limitation of applying dynamic models in planning is they are impractical to use such models for large networks.

Some mesoscopic simulation packages (e.g. SATURN) can be used to simulate short-term closures using the so-called "*quasi-dynamic*" principle (i.e. with residual queues). What the quasi-dynamic approach does is to make the traffic conditions at the end of a time-slice be the starting conditions for the subsequent time-slice (Van Vliet, 2015). Thus, if a network is modelled from 8:00am-10:00am, small intervals (10 minutes, say) could be used to estimate the flows and travel times in each interval. By using this feature, short-term degradation might be simulated using SATURN, since the traffic condition for short intervals can be captured. It should be noted that the duration of the interval is directly related to the simulation running time. For instance, for a two-hour simulation interval, using one minute time-slices will involve more running time than using 10 minute time-slices, as SATURN will analyse 120 and 12 scenarios, respectively. Overall, less detail and running time are needed compared with microscopic packages.

2.3. The CE method

The CE method, a Monte-Carlo method, was originally developed to estimate the probability of occurrence of rare events (e.g. the probability of failure of a particular network), then it was extended to solve combinatorial optimisation problems when the objective function is very complicated and it is necessary to do a lot of sampling. The reader may refer to a full description in Rubinstein and Kroese (2004).

The CE involves three main steps: generating a random sample from a pre-specified probability distribution function, evaluating the selected sample based on a performance index, then updating the samples based on a smoothing parameter (α) in which:

$$\gamma^{(t+1)} = \alpha\,\gamma_{new}{}^{(t)} + (1-\alpha)\gamma^{(t)} \tag{5}$$

where $\gamma^{(t+1)}$ is the set of parameter values that minimises $PI\,(\boldsymbol{X}, \boldsymbol{q_{UE}}\,\boldsymbol{(X)})$ in equation (1), t is the iteration number, $\gamma^{(t)}$ is the previous set of parameter values, and $\gamma_{new}{}^{(t)}$ is the new set of parameter values. Typically, the value of α varies between $0 < \alpha \leq 1$.

Each observation in this sample is scored for its performance as the solution to the specified optimisation problem. A fixed percentage of the best performing observations are referred to as the elite sample. The elite sample helps to update the parameters in the next generated solutions to improve the quality of the solution. This will be repeated until convergence occurs and an optimal solution is found.

To improve the algorithm's performance: 1) the population size can be increased to maximize the possibility of having a good random sample. 2) the smoothing factor (α) can be varied (it was found empirically that a value of α between 0.4 and 0.9 gives the best results (de Boer et al., 2005). 3) the percentage of the elite sample. It is worth mentioning that the process was repeated several times with different seed values for the random number generator to identify the robustness of the results (i.e. how sensitive they are to different seed values).

3. A study case to test the numerical model on a real network

The performance of the proposed approach was assessed by applying it to a real network.

3.1. The testbed description

The approach was tested on the Cambridge (UK) network (Fig. 3a), which comprises 141 zones, 1,091 links and 608 nodes, including 24 signalised junctions with 2-stage arrangements (Fig.3b). The common cycle length was fixed at 60 seconds, and all inter-greens were set to 5 seconds. The total demand in this network reflects one peak hour, with a total number of 42,023 commenced vehicle trips.

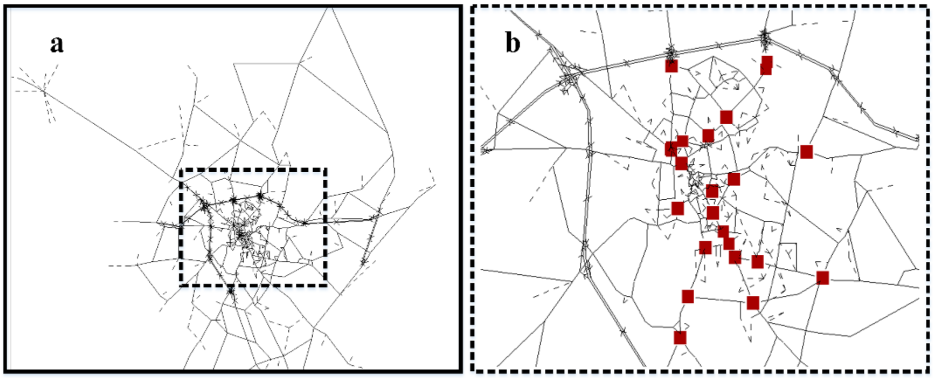

Fig. 3. (a) Cambridge network modelled in SATURN; (b) The 24 signalized intersections represented by red squares

The objective was to find the set of values for the 71 variables (i.e. 48 green times and 23 offsets) that minimises the travel time in the network in the case of disruption. These variables were constrained to be integers, with the minimum green times being set to 7 seconds, and the offsets ranging from zero up to 59 seconds, with the offset at node 2045 being zero. The traffic flow at the most congested intersection (node 2010) was degraded by applying several blockage scenarios; which involved various combinations of two factors (the duration and the % of capacity reduction of the blockage).

3.2. Simulation Results

The results of simulating different blockage scenarios (i.e. the green times and offsets) are summarised in Table 1 for node 2010 and the adjacent nodes 3089 and 2040. These results are for four levels of capacity reduction (0%, 25%, 50%, and 75%) at node 2010, for a period of one hour. Re-routing can result in changes to the optimal signal settings at the node where the disruption occurs and at other signalised intersections in the vicinity of the node where the disruption occurs. The results indicate that the optimal signal settings for node 2010 appear to be sensitive to the severity of the disruption. For instance, there is a 54% increase in the optimal green time at node 2010 with a 75% reduction in its capacity. Furthermore, the changes in the optimum green times and offsets for the nearby signalised intersections (i.e. nodes 3089 and 2040) show how the degradation at one node affects the optimal signal settings in the nearby nodes. Moreover, the changes as the capacity reduction increases from 0% to 75% are far from linear (i.e.

the optimal settings tend to fluctuate). For example, the offsets at 2010 are, respectively, 17s, 41s, 12s, and 18s and the green times for stage A at 2040 are, respectively, 43s, 22s, 43s, and 43s. SATURN captures the re-routing of drivers at the blocked node 2010 (Fig. 4).

Table 1. Green times (Stage A) and offsets for nodes: 2010, 3089, and 2040

Node / Reduction	At node 2010		At node 3089		At node 2040	
Capacity reduction	Green times (s)	Offsets(s)	Green times (s)	Offsets (s)	Green times (s)	Offsets (s)
0%	28	17	19	37	43	57
25%	23	41	22	37	22	9
50%	43	12	23	15	43	9
75%	43	18	12	28	43	9

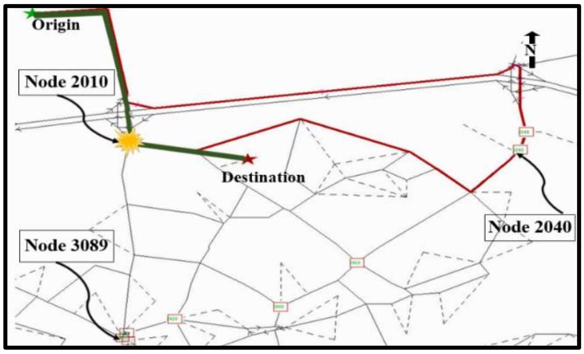

Fig.4. Optimum path obtained from SATURN without a blockage (green) and with a blockage (red)

The convergence of the green times and offsets for three nodes: 2010, 3089, and 2040 are presented in Fig. 5 and 6, respectively. This convergence is for 75 % reduction in capacity at node 2010. Initially, the probability of occurrence for each solution is equally likely with a probability of 0.0270 for each of the 37 possible green times and 0.0167 for each of the 60 possible offsets. The probability of each solution is then updated after each iteration based on the elite sample generated initially from a discrete uniform distribution with the mean and the standard deviation of the values in this sample to create a new distribution. The distribution becomes less uniform and more concentrated as the number of iterations increases, until the solution stabilises and has a probability close to one (the optimal value) of the variable. For instance, the probability of having a solution of 43s for the green time at node 2010 is one (Fig. 5c).

The convergence of offsets is difficult to obtain, particularly for the disrupted node 2010 (Fig. 6c). The offset values in Table 1 are those with the highest probability after 30 iterations. For example, for a 75% reduction in capacity at node 2010, the offset at node 2010 with the highest probability is 18s, which has a probability of 0.6. These values have been used for estimating the total travel times. One should note that the offsets at some nodes do not necessarily converge after 30 iterations, especially at nodes with closures (Fig.7). It can be seen that the convergence for green times is much quicker than for offsets. In the case of a 75% capacity reduction at node 2010 for one hour, the value of the objective function (i.e. the total travel time in the network) is decreased from 19,579 to 18,250 hours (i.e. about 7%). The mean of all possible solutions converges to a value of around 18,250 hours after 20 iterations.

The quasi-dynamic approach is used to simulate the traffic condition for short intervals, the simulated hour being divided into 15-minute intervals (i.e. 4 time slices) and 4-minute intervals (i.e. 15 time slices). The results obtained from applying the quasi-dynamic for different blockage duration (4, 12, and 20 minutes) are still under preparation.

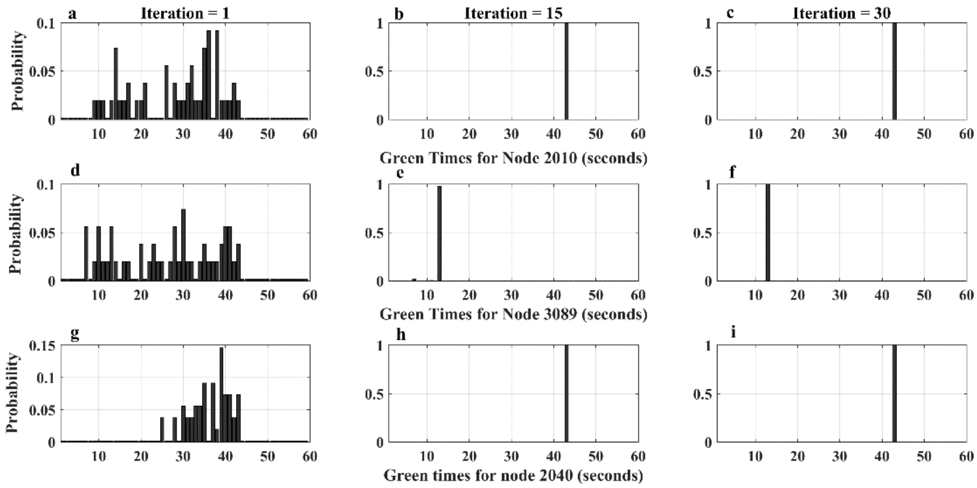

Fig.5. The convergence of green times, stage (A), in the case of 75% reduction for nodes: 2010 (a-c), 3089(d-f), and 2040(g-i)

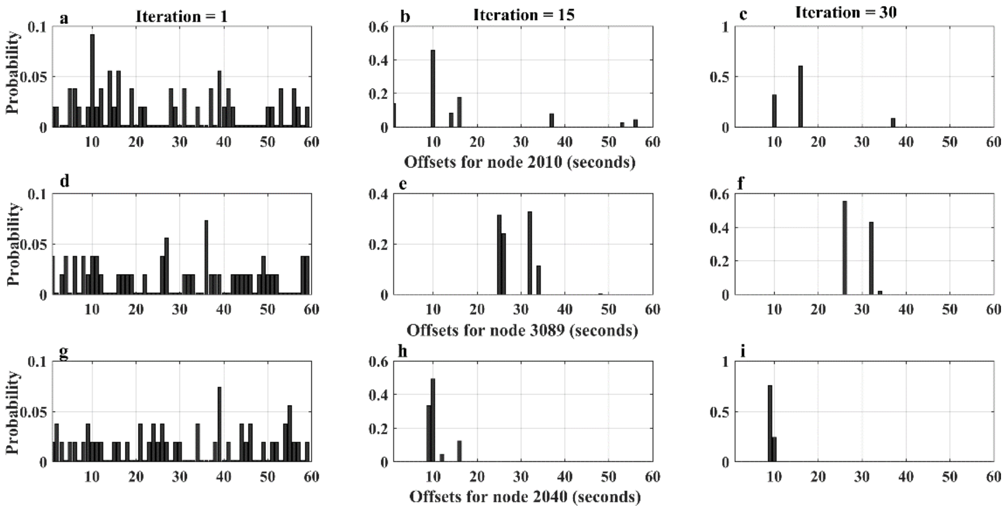

Fig.6. The convergence of the offsets in the case of 75% reduction for nodes: 2010 (a-c), 3089 (d-f), and 2040 (g-i)

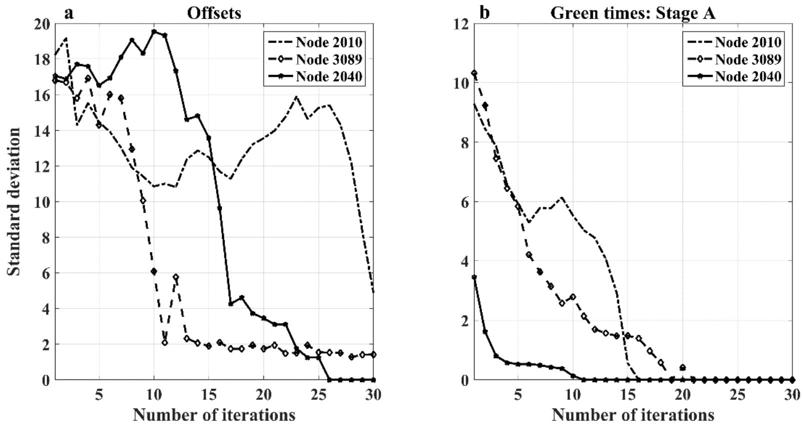

Fig.7. The standard deviation for nodes 2010, 3089, and 2040 in the case of 75% reduction in capacity for: a) the offsets; b) the green times

4. Conclusions and Recommendations

1. The results of the CE framework (as described above) indicates that there is a value in using the method to optimise traffic signal control to minimise the total travel time to assist traffic to divert around blockages. These results are for one hour closure, shorter time closures (e.g. 4 minutes) using the quasi-dynamic approach is still ongoing.

2. The most congested node was chosen as the location of a blockage, as network performance is expected to be sensitive to blockages at congested nodes. However, the exposure index introduced by Jenelius et al. (2006) might be a more appropriate indicator of network performance sensitivity and it is planned to investigate its use for identifying critical blockage locations.

3. The focus of this paper is on how we can assist road users to re-route to good alternative routes (i.e. to avoid the disrupted areas). Re-routing can result in changes to the optimal signal settings at the node where the disruption occurs and at other signalised intersections in the vicinity of the node where the disruption occurs.

References

ASBJORNSLETT, B. E. 1999. Assess the vulnerability of your production system. *Production Planning & Control,* 10**,** 219-229.

BERDICA, K. 2002. An introduction to road vulnerability: what has been done, is done and should be done. *Transport Policy,* 9**,** 117-127.

BRUNEAU, M., CHANG, S. E., EGUCHI, R. T., LEE, G. C., O'ROURKE, T. D., REINHORN, A. M., SHINOZUKA, M., TIERNEY, K., WALLACE, W. A. & VON WINTERFELDT, D. 2003. A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities. *Earthquake Spectra,* 19**,** 733-752.

DE BOER, P.-T., KROESE, D. P., MANNOR, S. & RUBINSTEIN, R. Y. 2005. A Tutorial on the Cross-Entropy Method. *Annals of Operations Research,* 134**,** 19-67.

JENELIUS, E., PETERSEN, T. & MATTSSON, L.-G. 2006. Importance and exposure in road network vulnerability analysis. *Transportation Research Part A: Policy and Practice,* 40**,** 537-560.

KAVIANI, A., THOMPSON, R. G. & RAJABIFARD, A. 2017. Improving regional road network resilience by optimised traffic guidance. *Transportmetrica A: Transport Science,* 13**,** 794-828.

KOOREY, G., MCMILLAN, S. & NICHOLSON, A. 2015. Incident Management and Network Performance. *Transportation Research Procedia,* 6**,** 3-16.Presented at the 4th ISTS conference.

MAHER, M. 2008. The Optimization of Signal Settings on a Signalized Roundabout Using the Cross-entropy Method. *Computer-Aided Civil and Infrastructure Engineering,* 23**,** 76-85.

MAHER, M., LIU, R. & NGODUY, D. 2013. Signal optimisation using the cross entropy method. *Transportation Research Part C: Emerging Technologies,* 27**,** 76-88.

MATTSSON, L.-G. & JENELIUS, E. 2015. Vulnerability and Resilience of Transport Systems – A Discussion of Recent Research. *Transportation Research Part A: Policy and Practice,* 81**,** 16-34.

NGODUY, D. & MAHER, M.2011. Cross Entropy Method for a Deterministic Optimal Signalization in an Urban Network. Transportation Research Board 90th Annual MeetingTransportation Research Board.

NGODUY, D. & MAHER, M. J. 2012. Calibration of second order traffic models using continuous cross entropy method. *Transportation Research Part C: Emerging Technologies,* 24**,** 102-121.

PEARCE, V. 2000. Incident management successful practices: a cross-cutting study: improving mobility and saving lives. United States. Joint Program Office for Intelligent Transportation Systems.

RUBINSTEIN, R. Y. 1997. Optimization of computer simulation models with rare events. *European Journal of Operational Research,* 99**,** 89-112.

RUBINSTEIN, R. Y. & KROESE, D. P. 2004. *The Cross-Entropy Method,* New York, NY, Springer New York.

SCHRANK, D., LOMAX, T. & TURNER, S. 2009. Urban mobility report texas transportation institute. *Texas: Texas Transportation Institute*.

SHEFFI, Y. R. J. 2005. *The Resilient Enterprise: Overcoming Vulnerability for Competitive Advantage,* USA, MIT Press.

TAYLOR, M. 2017. *Vulnerability analysis for transportation networks*, Elsevier.

VAN VLIET, D. 2015. *SATURN Manual- Version 11.3,* ATKINS.

ZHONG, R. X., FU, K. Y., SUMALEE, A., NGODUY, D. & LAM, W. H. K. 2016. A cross-entropy method and probabilistic sensitivity analysis framework for calibrating microscopic traffic models. *Transportation Research Part C: Emerging Technologies,* 63**,** 147-169.