Applying Mesoscopic Simulation to Evacuation Planning for the Houston Region

Chi Ping Lam, Houston-Galveston Area Council

Colby Brown, Citilabs

Chris Van Slyke, Houston-Galveston Area Council

Heng Wang, Houston-Galveston Area Council

Abstract

In September 2005, over one million residents evacuated from the greater Houston area in response to Hurricane Rita, causing severe congestion on the region's highways. In recognition of the consequences of this event, the Houston-Galveston Area Council (H-GAC) identified the need for more efficient evacuation plan coordination. To aid in this effort H-GAC worked with Texas Transportation Institute (TTI) and Catalans to develop an evacuation model capable of replicating the Rita event and evaluating future evacuation scenarios.

The evacuation model simulates two different kinds of traffic simultaneously: the evacuation and background travel demands. Evacuation traffic, with limited knowledge of alternatives, concentrates on evacuation routes and freeways despite the existence of faster paths, while background traffic follows normal daily travel patterns. Dynamic mesoscopic traffic assignment was adopted because it could simulate queuing, estimating travel time delay and clearance time impacts within the evacuation period without requiring any significant increase in network detail.

However, with 3000 zones, over 10 million regular trips and 1 million plus evacuating trips, the enormous size of the H-GAC network presented a major challenge for mesoscopic simulation. To reduce running time and lessen memory consumption, probabilistic trip integerization and zonal aggregation techniques were applied. The network was also reviewed and modified to eliminate artificial congestion due to incorrect network coding. Finally, various assignment methodologies were tested for best performance in both running time and memory consumption.

Acknowledgements

We could like to specially thank Alan Clark, the executive director of Houston-Galveston Area Council, for his continuous support for this project. We could also like to thank Matthew Martimo as he worked hard to improve the computer application and optimize its running time. He also delicately testing the model and implements the original randomly trip distribution by time. We could like to acknowledge other contributors, such as Dr. Jim Benson, Andy Mullins, Philip Reeder, David Schellinger, Wade White, Vladimir Magana, Pete Mazurka, Ken Vaughn, and others not named here.

Goal of the Model

The purpose of the evacuation model is to numerically evaluate different evacuation scenarios given various demands and strategies. The evacuation criteria are the delay, length of queuing, travel time to exit the region, and through volumes getting out the region. The model should have the ability in model common evacuation strategies, including managed departure and contra-flow lanes.

The model will be validated through the normal day (no evacuation) scenario and the Rita scenario. The validation is to prove that the modal is able to reflect reality.

The Case for Mesoscopic Assignment
The H-GAC regional travel demand model employs tradition macroscopic assignment to forecast travel demands. However, the macroscopic assignment is not suitable for this project because it cannot model queuing and the dynamic nature of evacuation. Usually micro-simulation is applied in dynamic traffic assignment, but it is infeasible for this large and complex study area. First, most simulation packages cannot simulate over 30 million trips occurs in the evacuation. Second, micro-simulation requires accurate and detail information on every road segments and intersection, and it is manually impossible to collect and to code on the regional network.

Mesoscopic assignment methodology, a bridge connecting macroscopic and microscopic assignments, is better suited for this model. It is a flow-based algorithm. It calculates travel time by volume-delay function and queuing algorithm. If the traffic demand is below the capacity and the downstream link is clear, the travel time is determined solely by volume-delay function. If the incoming vehicle demand exceeds the capacity, the overflow vehicles are blocked at the upstream links and wait until the downstream vehicles have been dispatched. By enforcing the capacity limit, it estimates more realistic travel time, delay and volumes than the static assignment. Many mesoscopic computer applications also accept intersection configurations and controls but this is not necessary. On the micro level, mesoscopic assignment tracks individual vehicle packet over time. It preserves the experience of every individual packet. It is designed to quickly analyze a large area with the details of dynamic queuing and to capture individual travel experience – the two major criteria to evaluate evacuation scenarios. Avenue, a Citilab's add-on product to perform mesoscopic assignment, is used in this model.

Performance Issues
The initial model, without any calibration, is subjected to many performance issues. It often crashes due to "out of memory error." Occasionally the model completes but it took a few days! Worse, the results do not make sense and the congestion appears at the wrong location at the wrong time! There are four major causes for all these performance issues. The causes include:

- The 3000 zone structure is large that takes very long time for path building.
- High number of packets on the network.
- Some routes were overloaded due to insufficient number of iterations and path choices.
- Artificial congestion due to network and demand coding errors.

The number of individual packets is an important factor determining the model performance. A packet is a group of vehicles traveling from the same origin zone to the same destination zone in the same time segment. In theory, packet can hold any number of vehicle trips; but in the model each packet is restricted to hold maximum 10 to 20 vehicle trips.

This memory problem is caused by the memory constraints posted by Windows XP. Cube Avenue could track 6 million packets under the 2GB. Memory is consumed to track movements of packets currently traveling on the network, and any packet which has not start its trip or has reached its destination does not consume memory. If each packet could hold maximum 20 packets in the model, the software could track at most 120 million vehicles at any time, which is far more than the 30 million trips for the entire evacuation periods. It is logical to conclude that Cube Avenue is capable to simulate the evacuation if the computer resources are used more effectively.

The following sections discuss strategies to decrease the number of package, and other methods to improve the performance.

Probabilistic Hourly Trip Table

The evacuation model simulates the traffic based on hourly trip tables. The traditional travel demand model calculates the hourly demands by multiplying the demands of entire modeling period to fractional time-of-day factor. Many packets hold small number of daily trips. Next, multiplying the daily trips to the hourly factor, those trips are further spilt to 24 smaller hourly packets per day. Many of these hourly packets hold less than 1 vehicle trip. A packet, no matter whether it holds 0.01 vehicles to 9.9 vehicles, consumes same amount of memory. Consider the following example in table 1. This zonal interchange has 12.93 daily HOV toll trips. The second and the third columns in figure 1 are the hourly trips after multiplying the hourly factor, and the fourth column is the sum of second and third columns. In this example, 21 hourly packets in the fourth column have less than 1 vehicle trips!

The number of hourly fractional trip packet can be reduced by treating the hourly factor as a probabilistic function. In reality, it is impossible to divide a vehicle to fractional vehicle trips. Therefore, the hourly factor could be viewed as the probabilistic distribution of vehicle departure time. The model then randomly distributes vehicle time over time according to this distribution function. Back to the example in figure 1, this methodology splits the 12.93 daily trips to 12 packets. There is one single trip for the first 11 packets, and there are 1.93 trips in the last packet to preserve the total number of daily trips. The fourth column is now the probability for a single packet to depart in a particular hour. The fifth column is the randomly assigned hourly trips.

The probabilistic hourly trip distribution successfully reduces the number of packets. It reduces 95.7% of regular day trip packets within 24-hour period and reduces 98.2% of the evacuation trip packets within the 72-hour period.

Zonal Aggregation

The number of zone is another factor affecting the amount of running time and memory consumption. Experimental testing shows that Avenue could spend more than one-half of total running time on path building. If the zones are aggregated, the total path building time could be significantly decreased.

	Interchange 469 to 2398			
HR	2 Person Pay	3+ Person Pay	Total	Assign Integer Trips
1	0.0038	0.0063	0.01	0
2	0.0026	0.0048	0.01	0
3	0.0026	0.0048	0.01	0
4	0.0078	0.0072	0.02	1
5	0.0113	0.0076	0.02	0
6	0.039	0.0197	0.06	0
7	0.1438	0.0947	0.24	0
8	0.2921	0.2983	0.59	1
9	0.2151	0.2103	0.43	0
10	0.1493	0.1638	0.31	0
11	0.2423	0.2571	0.50	1
12	0.2502	0.2993	0.55	0
13	0.3012	0.367	0.67	1
14	0.2948	0.3224	0.62	0
15	0.4754	0.3876	0.86	1
16	0.8557	0.6355	1.49	1.93
17	0.7165	0.5194	1.24	1
18	0.728	0.4955	1.22	1
19	0.598	0.3964	0.99	1
20	0.4641	0.2981	0.76	1
21	0.5091	0.3129	0.82	1
22	0.4972	0.2932	0.79	0
23	0.3251	0.1853	0.51	1
24	0.1339	0.0808	0.21	0
Daily	7.2589	5.668	**12.93**	**12.93**

Figure 1 Hourly Trip by Multiplying Hourly Factor

The TAZs are aggregated according to principal arterials, natural barriers, and size. Generally, the aggregated zonal boundary consists of natural barrier, freeway, major evacuation routes, or principal arterials. The aggregated zone could cross over the minor arterials and local collectors. The natural barriers considered include Galveston Bay and major water streams. The size of the TAZ should not be too large. There are a few aggregated zones produce or attract more than eight-thousand trips in a single hour. Most of these aggregated zones are split to smaller zones to avoid uneven or unrealistic loadings.

After the zonal aggregation, the number of zones reduced from 3000 zones to around 570 zones. The model running time had been decreased from weeks to at most two days, and the memory problem rarely occurs. The model is now ready for calibration.

Number of Iterations and Maximum Path Choices

In mesoscopic assignment, in early iteration, when many packets are trying to pack into the shortest path, many packets are forced to wait patiently on upstream links and are not clear even after a long time. Insufficient number of iterations will create artificial congestion. Matthew Martimo has demonstrated the impact of insufficient number of iterations though an hourly mesoscopic assignment. The test loads the 7am hourly demand in the first hour of simulation, and then runs the second hour of simulation without any additional demand. As illustrated in figure 2, for the first iteration, there are 89,000 packets remaining stuck in the network at the end of second hour. The number of packets remaining reduced to 60,000 in iteration 2, and it is eventually reduced to only 2,000 in iterations 10. Despite supply and demands are identical for every iterations, the congestion is far worse in iteration 1. Sufficient number of iteration is required to distribute the normal day traffic evenly.

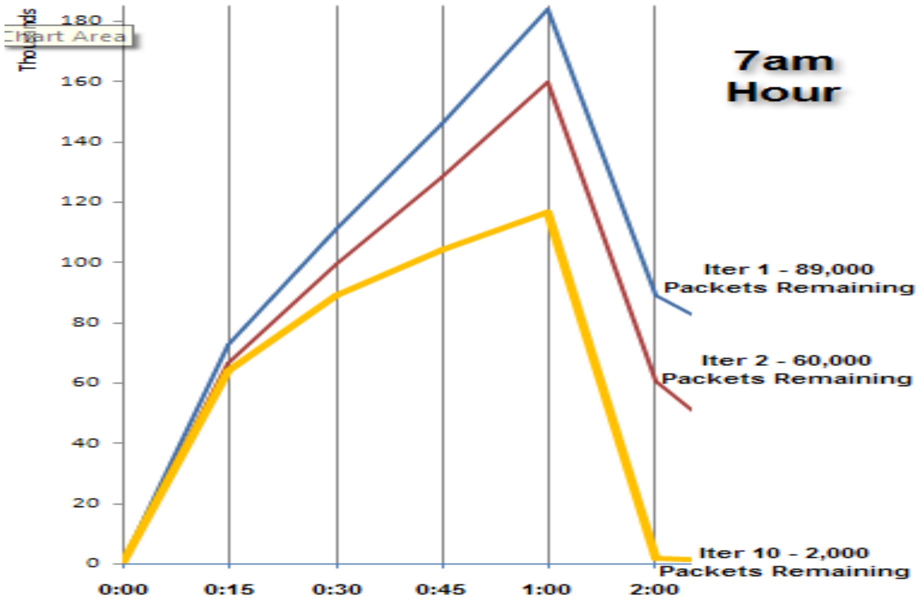

Figure 2 Assign 7 AM hourly demands with different number of iterations

Similarly, the number of path choices also affects the level of congestion. Limiting the number of path choices is a method to reduce the path-building time; but restricting it too much could create artificial congestion by cramming many packets to the few routes.

We have tried various numbers of iterations and path choices on the model. There are eight iterations and maximum four path choices in the version of the model submitted by Citilabs, and latest model has 30 iterations and maximum 12 path choices.

Adjusting BPR Curves

The evacuation model initially borrows the standard BPR curve from the macroscopic model, but the parameters of the BPR curves have been modified to fit the mesoscopic assignment. The traditional BPR curve, with alpha value 0.15 and beta value 4, is intended for an assignment begins with free flow speed and design capacity, and the assigned volume is allowed to exceed the design capacity. The design capacity is generally referred as the Level-Of-Service (LOS) C capacity, which is less than the maximum capacity. The alpha value 0.15 determines that when volume equals to the capacity, the speed is 15% less of the free flow speed. The H-GAC macroscopic assignment modifies this presumption by using the LOS C speed and maximum capacity in the standard BPR function in order to forecast more accurate travel demand.

Dynamic traffic assignment requires a more precise speed-capacity relationship, and the standard BPR function fails to fit the need. The evacuation model is supposed to quantify the travel time and delay correctly. Using LOS C speed as the free flow speed could over-estimate travel time and under-forecast freeway volume in free-flow condition. Therefore, to correctly estimate travel time and delay, the evacuation model must use the true free flow speed, instead of the LOS C speed, as the input to the volume-delay function. In this case, the standard BPR function could over-estimate the speed at maximum capacity.

We pick two set of BPR parameters for freeway and local arterials. For freeway, the alpha is 1 and the beta is 3. For local arterials, the alpha is 1 and the beta is 2. Our BPR curves, along with

Horowitz curve and the standard BPR curves, are plotted in figure 3 for a segment with 65mph free flow speed. The speed deteriorates faster in the local arterial curve than in the freeway curve. Overall, our BPR functions reduce speeds quicker in small volume-capacity ratio; but in reality, particularly on freeway, the speeds stay relatively constant on low volume. Our BPR functions exaggerate speed deterioration on low volume-capacity ratio in order to force the path-finder be sensitive to volume changes.

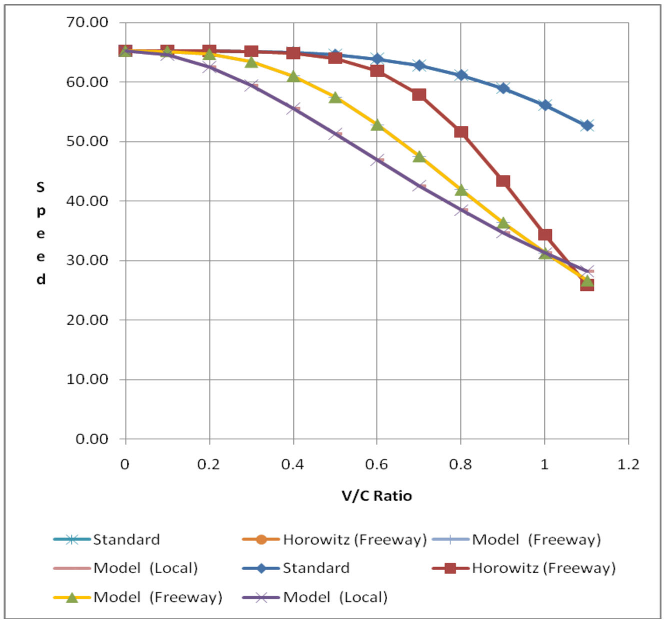

Figure 3 Various BPR Curves

Overview of Validation: Re-Generate the Real World Scenarios

The first scenario, the regular day scenario, is the no-evacuation scenario. Evacuation begins in a normal day when everyone is doing their normal business. In the Rita evacuation, some coastal residents evacuated early but most people are continuing their regular routines. Therefore, the early evacuation traffic interacts heavily with the regular day traffic, and both traffic impacts the roadway condition. Even in late evacuation period, there are still a few people, particularly residents living far away from the coast, continue their routine business. The evacuation model must be able to forecast regular day traffic in order to simulate the interaction between regular day traffic and evacuation traffic.

The evacuation model is also expected to re-generate the Rita scenario. If the evacuation model could not re-generate the Rita scenario, how could anyone believe that it could forecast evacuation accurately? However, unlike the regular day scenario, it is very difficult to validate the Rita scenario due to lack of information. There are very few traffic count data available. There are also some speed data collected by the Automatic Vehicle Identification (AVI) technology, but those AVI data is confined with Harris County. The evacuation model should simulate comparable speeds and volumes at locations where traffic counts or AVI data are available, and produces some logical and reasonable speed and volumes at area without any available data.

Validate Regular Day Scenario

For the regular day scenario, or the no evacuation scenario, only the regular day traffic is loaded into the model. It simulates a twenty-hour period because the regional model assumes same daily trip table for every weekday. The daily trip table is split to 24 hourly trip tables by the probabilistic methodology describe above.

Many network coding problems are discovered through the validation. The regional travel demand model allows links with volume-capacity over one to indicate segments with high demands, even though by definition volume should never be able to exceed capacity. However, for some links, the high volume-capacity ratio may be caused by network coding issues. The following network coding missteps could cause unusually high volume-capacity ratio:

1. Insufficient number of centroid connectors
2. Incorrect number of lanes or facility types is entered.
3. The road has been expanded but the expansion is not updated in the regional network.
4. Turning lanes and auxiliary lanes are not coded in the regional network at all.

These network coding issues are “hidden” because those issues are disguised as high demand links as both cases result in high volume over capacity ratio. However, in mesoscopic assignment, those issues are exposed because it does not allow volume over capacity. The vehicles are blocked at the upstream links and, in turn, create unexpected congestion.

In all the coding issues, turning lanes and auxiliary lanes are systematic problems. The travel demand model does not include turning lanes and auxiliary lanes as a policy, because those lanes do not provide continuous through capacity over a long segment. However, for dynamic traffic assignment, those lanes are very critical because they supply crucial capacity for vehicles getting in and out major intersections or interchanges, as well as provide additional physical space to store queued vehicles. The missing auxiliary lanes exaggerate congestion at several locations, and the most severe location is the Galleria area.

The Galleria area is an major activity center. At the southeast corner of Galleria is the major change of US 59 at Interstate 610, an interchange known for its traffic jam. The area is very congested in peak periods, but the model predicts heavy congestion at 10 pm! In figure 4, the red color links are road with average speed less than 10 mph between 8-9am. The majority of road links are project to below 10 mph between 8 to 9 pm; this level of congestion is too high even in this congested area. Moreover, the queues back up several miles from the freeway interchange, spreading the congestion to a very large area. The demand for this area seems to be in the range; all the evidence suggests that the there are enough capacity codes on the network.

Checking with the aerial photo and Google Earth, we found that all the auxiliary lanes at the interchanges are missing. There are two to three auxiliary lanes along with three to four main lanes at each direction. The network also misses the fly-over ramps where traffic could pass over the interchange to directly access or egress to Galleria. Figure 4 illustrates all the road segments with missing lanes added.

After adding all the missing auxiliary lanes and fly-over ramps, the congestion disappears at 9 pm, as shown in figure 6. There are some slow speed links remain on busy intersections or on locations of uneven centroid loadings, but the congestion does not spread to a large area. We have also reviewed other locations with unusual congestion assigned, and corrected the underneath network or demand issues.

Figure 4 Slow Roads at Galleria Area between 8-9 PM before Adding Auxiliary Lanes

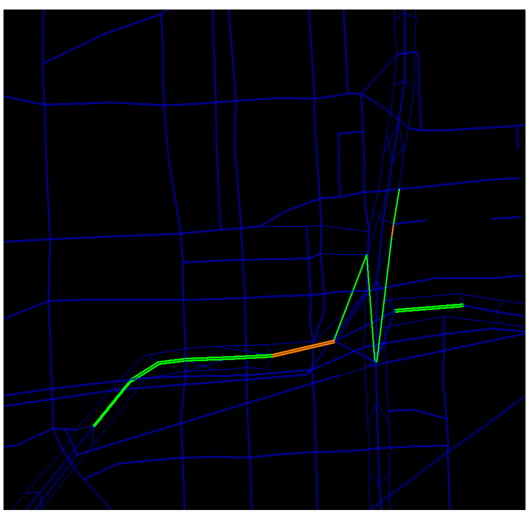

Figure 5 Lanes Added

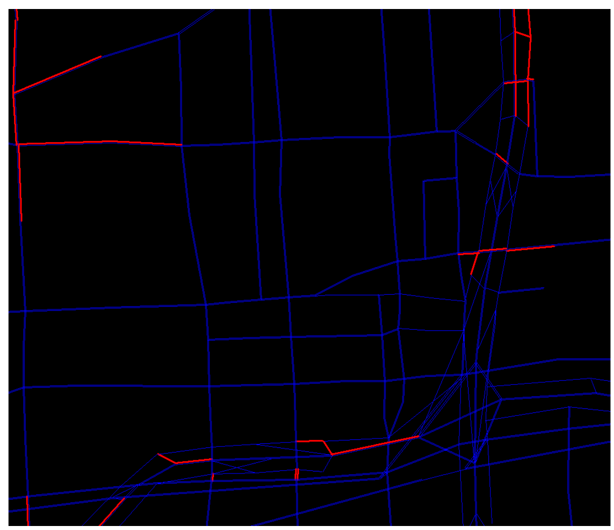

Figure 6 Slow Roads After adding Auxiliary Lanes

Another validation technique is comparing the vehicle miles total between the macroscopic assignment and the mesoscopic assignment. For consistency issues, the same trip tables, network, and zone structure are used in both assignment algorithms. The BPR parameters are different for both assignments. It is encouraging that there is only 1% difference between the two total VMT, suggesting the mesoscopic algorithm assigns reasonable routes.

The future validation step is to compare the assigned volumes with traffic count. Due to the aggregated zone structure, the volumes on individual links could be significantly different than the count. Yet, the total screen line volume assigned should be within acceptable range of the screen line count.

Validate Rita Event

The Rita event was a six-day evacuation before the hurricane lands. Out of the six days, the three consecutive days with over 90% of evacuation trips are selected to be simulated in the evacuation model. In the first one and half days out of these three days, most evacuees depart from the mandatory zones, and the congestion is relatively less serious. Most residents living in non-coastal area continue to work and to do normal business. In the next 1.5 days, the evacuation becomes region wide. Schools and works are cancel, and people evacuates from everywhere in the region. The congestion is much more severe and spread over entire region, and many evacuation routes and freeway corridors become gridlock.

There are three main trip purposes in the Rita evacuation. The first trip purpose, obviously, is the evacuation trips that intend to stay in another place where is perceived to be safer than home. More than 90% of the evacuation trips leave the region. Many evacuees stuck on evacuation routes, major highway, or freeway because they do not possess enough knowledge on local routes outside their local areas in order to seek an alternative route on local arterials. It is also reported that some exit ramps are blocked to keep the evacuation traffic staying of freeway. Their travel behavior is somewhat similar to all-or-nothing assignment nature.

The second trip purpose is the regular day trips that people made daily. The regular daily trips behave in user-equilibrium nature, and the drivers should be able to find an alternative routes if their routine path are blocked by evacuation traffic. As the evacuation continues, there are less and less people to make their regular daily trips. A set of hourly reduction factors are develop to reduce the number of regular daily trips as the evacuation going on.

The third trip purpose is called non-evacuate special trips. These trips are made by people decide to stay in the region. A majority of these trips are made by non-evacuating residents preparing for the upcoming disaster. For example, they buy woods and nails in hardware stores, fill up the tanks in gasoline stations, and purchase water and non-perishable food in grocery stores.

The evacuation model must assign all three purposes of trips simultaneously. The regular daily trips and the non-evacuate special trips are combined into the same class because both trips are user-equilibrium nature. Another challenge is how to assign the evacuation trips. In the model, those trips received less number of iterations than the regular daily trips and non-evacuate special trips to reflect the fact that evacuation trips are less robust. A local deter factor is also introduced to mimic the evacuee's unfamiliarity to local arterials; it multiplies the travel time on every local streets in their cost functions, discouraging the evacuation trips to be assigned on local routes. The local deter factor does not apply to the regular daily trips and non-evacuate special trips.

Comparing with the AVI speed data, the model seems to predict more severe congestion than the AVI data indicated. It is encouraging as the model exhibits ability to create the gridlock condition. We will continue to work on validating Rita scenario.

Goals after the Validation

Simulating the regular day scenario or the Rita event is not the ultimate purpose of this project. The evacuation model is intended to be a tool evaluating different evacuation strategies on both supply and demand sides. We will run a series of common strategies on the evacuation model to test its sensitivity to these strategies. These strategies include contra flow lanes, managed Departure , and additional evacuation routes on local arterials

Lessons Learned

The most challenging aspect of this project is for modelers to be comfortable with mesoscopic model. The traditional travel demand modelers wonder if the mesoscopic model is too sensitive to capacity and doubt if it is robust enough to deal with future demands. The question could be partially answered by trying a more robust assignment algorithm with sufficient number of iterations and path choices, along with rigorous network checking. Some travel demand modelers also feel scary about randomness the evacuation model because they are used to the stable, static macroscopic assignment. On the other hand, the micro-simulation engineers argue that capacity is a fluctuating variable instead of an absolute number. Even if a link is loaded with demands over the capacity, the vehicles should still be moving in and out the queue slowly. Although their understanding on capacity is correct in microscopic level, they do not realize that capacity is here to simplify network coding and to reduce complexity of assignment. Overall, there is a period of confusion and argument before we accepted the strength and weakness of mesoscopic assignment.

The second lesson we learned is that good data preparation is essential for mesoscopic assignment. This project could be more efficient if we have reviewed our network data before the validation, saving much time of checking the network coding after many model runs are already made in the validation phase. The following suggestions could help preparing input data for mesoscopic assignment:

1. Check the entire network coding, especially on congested area and bottlenecks. Make sure the auxiliary lanes and turn lanes are included in the network coding.

2. Identify areas in the static model where the demands are over supply

3. Collect hourly traffic count or survey data for creating or to updating hourly split factors.

4. Collect hourly traffic count and speed data for validation.

5. Collect signal information such as optimization or synchronization plan.

The third lesson we learn is nothing new: starts with something small before going big. If there is a mulligan, we could start this project in a sub-area. There are too many hidden network and demand issues in the regional network, and it is difficult to diagnose these local issues in the big regional network, despite these local issues may impact a large area. A sub-area model is more efficient to identity local network or demand issues.

We recommend calibrating the mesoscopic model on a small, congested area, such as the CBD. Correct any network or demand issue in the congested area is critical because its impact will spread regionally. Perform multiple sensitivity tests in this sub-area model to learn how the mesoscopic model reacts with changes. If there are multiple congested areas in the regional network, run a sub-area mesoscopic assignment on each congested areas to identify and to correct all issues on those areas. Finally, validate the regional mesoscopic model with the calibrated parameters, cleaned networks, and the adjusted demands from the subarea assignments.