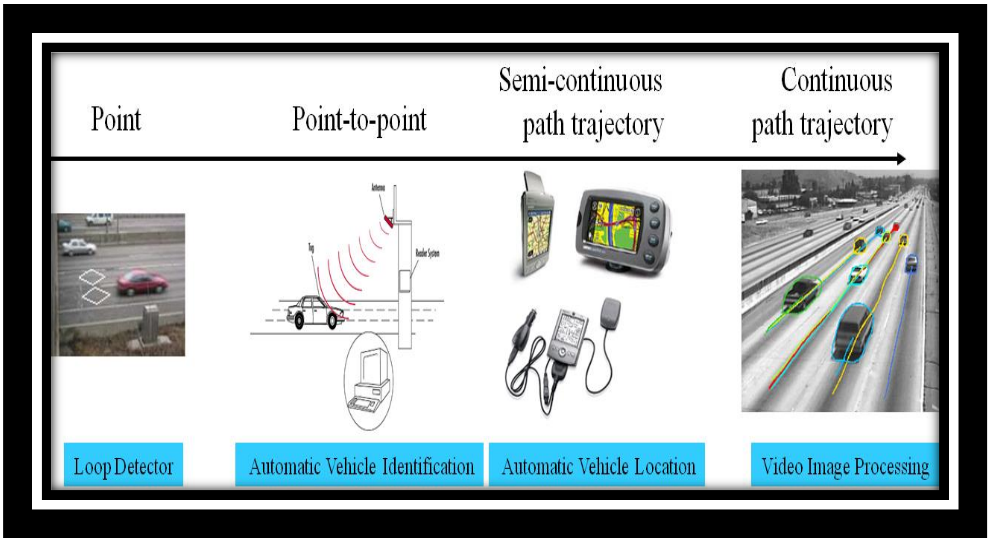
Lesson 5: Origin Destination Demand Matrix Estimation (ODME)

Version 1: Prepared by Xuesong Zhou (*xzhou99@gmail.com*)

Version 2: Prepared by Dana Alsader (dana.alsader@asu.edu)



Learning Objectives:

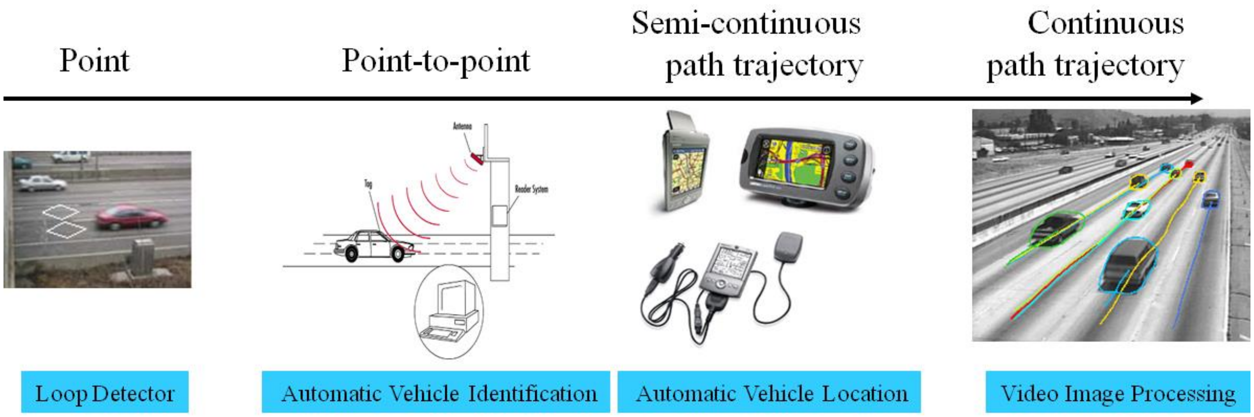
1. Understand the three important components for network OD demand matrix estimation
    a. Input data include link flow, turning movement flow counts, target/historical OD demand matrix
    b. Mapping matrix between OD demand and link/movement flow counts
    c. Mapping from all related OD flows to the link/movement volume of interest
2. Use a transportation network with link counts and movement turning counts to construct an optimization model with unknown OD demand matrix
3. Understand mathematical programming models with nonlinear objective function related to deviations with respect observations and target matrix .
4. Solve a real-life ODME problems using Excel Solver and a manual preparation process based on entry flow counts and turning percentage
5. Understand the basic concept of model validation using estimated and observed link counts through X/Y scatter plot
6. Visualize OD demand volume and link volume using 3-D interface through Google Earth

1. Introduction

Accurate origin-destination (OD) trip volume estimates are required by the many traffic planning applications to evaluate network flow conditions that result from the travel decisions of individual travelers/agents. Moreover, on-line applications of intelligent traffic network management call for the reliable forecasts of dynamic origin-destination demand and resulting network flow states so that proactive, coordinated traffic information and route guidance instructions can be generated to network travelers for their pre-trip planning and en-route diversion.

However, estimating a spatial traffic demand matrix is difficult in its own right, as traffic demand can vary significantly by time of day and day of week over different locations and evolve dramatically due to the feedback of implemented strategies. The inability of providing high quality OD demand estimates becomes a critical bottleneck in the evaluation and implementation of various traffic information and management scenarios, and consequently limits the potential for Intelligent Transportation Systems (ITS) deployments to alleviate traffic congestion and enhance mobility in urban networks.

There are a number of surveillance techniques available for the traffic monitoring and management purposes, each with ability to collect and process real-time traffic data in specific types, including point, point-to-point and path measurements.



[Point Sensors] By counting traffic passing through a specific location during a period of time, a wide range of vehicle detection devices provide various point measurements such as lane occupancy, traffic volume, vehicle headway, as well as time-mean speed. As the earliest vehicle detection device, pneumatic tubes have been applied in traffic engineering practice since the 1930s, and they are still commonly in use as temporary counting devices. In the 1960s, many intrusive sensors such as inductive loops, magnetometers, and piezoelectric cables were introduced for automatic vehicle detection and classification. Among them, inductive loops have become the predominant vehicle detection device in the United States, due to their associated low unit equipment cost and relatively high performance. On the other hand, intrusive type detectors have to be directly installed on the pavement surface, causing considerable traffic disruption and high risk for maintenance workers during the installation and repair activities. High failure rates and significant downtimes are two other major issues in operating inductive loop detectors. For example, Bikowitz and Ross (1982) indicated that approximately 25 percent of inductive loop detectors in New York State were not functioning properly at any given time. To overcome the disadvantages of inductive loop detectors, many roadside and overhead sensors are developed, including passive acoustic, passive infrared, and microwave radar detectors. These non-intrusive devices are able to provide traffic measures without stopping traffic in installation and maintenance.

[Vehicle Identification Devices] Many vehicle identification devices have been developed to track the identities of vehicles through mounted transponder tags or license plate numbers when vehicles pass multiple but non-contiguous reader stations. A raw tag read typically records a vehicle ID number, the related time stamp and passing site location. If two readers at different locations sequentially identify the same probe vehicle, then the corresponding data reads can be fused to calculate the journey travel time and the counts of identified vehicles between instrumented points.

[License Plate Surveys] In conventional license plate surveys, part of a registration number (e.g. only last three digits) might be recorded in order to reduce manual data collection effort and avoid high recording errors when recording the complete registration number. Several statistical methods (e.g. Makowski and Sinha, 1976; Maher, 1985; Watling and Maher, 1992) have been presented to reduce "spurious matches", which indicates that different vehicles observed at two points share the same partial registration number.

Automatic license plate matching techniques have entered the traffic surveillance field since 1970s, and many statistical and heuristic methods have been proposed to reduce reading errors and to provide reliable data association (Turner et al., 1998). Due to the difficulties in recognizing dirty and obscure characters, license plate based vehicle identification techniques typically lead to relatively low identification rates.

Many feature-based vision and pattern recognition algorithms (e.g. Evans, 1993; Shuldiner et al., 1996; Coifman et al., 1998) have been presented to track individual vehicle trajectories using camera surveillance data. By means of vehicle signature matching techniques (Coifman, 1998), several coupled point detectors can also be used to approximate point-to-point travel measures such as link segment travel time information. However, the required high coverage densities for vehicle signature matching techniques dramatically reduce the economic feasibility of their application in a large-scale network.

[Automatic Vehicle Identification Data] Radio Frequency Identification (RFID) technologies first appeared in AVI applications during the 1980s and has become a mature traffic surveillance technology that produces various traffic measures with high accuracy and reliability. Currently, many RFID-based AVI systems are widely deployed in road pricing, parking lot management, as well as real-time travel time information provision. For instance, around 51 AVI sites were installed and approximately 48,000 tags had been distributed to users at San Antonio by 2001, corresponding to a 5% market penetration rate, while Houston's TranStar fully relies on AVI data to provide travel time information currently (Haas et al. 2001). It should be also noted that, utilizing AVI data for traffic OD volume estimation, especially in the early deployment stage, can be constrained by low market penetration rates of AVI tags. A simulation-based study conducted by Van Aerde et al. (1993) shows that low market penetration rates directly result in small data samples in statistical reference and high variances in travel time and OD flow estimates.

[Automatic Vehicle Location data] With advances in Geographic Information System (GIS) and telecommunication, many automatic vehicle location (AVL) technologies, such as Global Positioning System (GPS), electronic distance measuring instruments (DMI's) and cellular telephone tracking, provide new possibilities for traffic monitoring to semi-continuously obtain detailed passing time and location information along individual vehicle trajectories. As pointed out by Tavana et al. (1999), the popular use of cellular phones can dramatically increase the quality and quantity of traffic data, as a source of probe vehicle information as well as a source of live human reports. However, privacy concerns and expensive one-time installation costs are two important disadvantages influencing the AVL deployment progress.

2. Mathematical background

2.1 OD demand estimation formulation

The general OD demand estimation problem aims to find an estimate of OD demand matrix by effectively utilizing traffic flow observations and other available information. Existing OD demand estimation models belong to two major categories: static models and dynamic models. Assuming constant trip desires over the estimation horizon, **static OD demand estimation models** estimate a static OD demand table based on daily or hourly average traffic counts. To realistically represent traffic formation and congestion on the traffic network, **dynamic OD demand estimation models** utilize time-varying traffic flow observations to estimate traffic demand that varies over time.

Notation:

l, s = link index in a transportation network

C =the vector of estimated link counts

C' =the vector of observed link counts,

D = the OD demand matrix $[d_{i,j}]$ from origin i to destination j, where $d_{i,j} \geq 0$.

D' = target or historical OD matrix $[d'_{i,j}]$ from origin i to destination j

A(D) = traffic assignment function that provides a mapping matrix between OD flows and link flows,

$p_{l,(i,j)}$ = Link flow proportion matrix, which describes the fraction of vehicular demand flow from **OD pair(*i, j*)** contributing to the flow on **link *l***

$p_{l,s,(i,j)}$ = Movement flow proportion matrix, which describes the fraction of vehicular demand flow from **OD pair(*i, j*)** contributing to the movement counts from **link *l*** to **link *s***

L = objective function, which can be expressed as a transformed function of likelihood measure

[Poisson distribution]

Early studies (e.g. Van Zuylen and Branston, 1982) assumed that averaged link counts follow a Poisson distribution, and set up the log likelihood function (1) subject to link volume observation constraint (2), and then estimated the OD demand matrix through the Maximum Likelihood. The assignment mapping matrix can be obtained by combining route choice proportions and the link path incidence matrix.

$$L = \text{Sum over } [C' \ln A(D) - A(D)] + \text{constant} \qquad (1)$$

$$\text{Subject to } C' = A(D) \qquad (2)$$

$$d_{i,j} \geq 0.$$

[Multivariate Normal Distribution]

By assuming a multivariate Normal distribution for traffic counts, Maher (1983) and Cascetta (1984) proposed a Bayesian estimator and a generalized least squares (GLS) estimator, respectively. If the error terms are Normal variables with zero mean and variance-covariance matrix W, the corresponding log likelihood function is

$$Ln\ L = -1/2\ [C' - A(D)]^T W^{-1} [C' - A(D)] + \text{constant}. \quad (3)$$

In a real traffic network, the number of links with observations is more likely to be less than the number of unknown OD pairs, and static OD demand estimation that purely relies on averaged link counts might lead to an underdetermined system. As a result, additional information must be supplemented to find a unique OD demand estimate.

[With *A priori* Information]

A priori information on trip demand, which typically comes from either sample surveys or outdated estimates, has been widely used as an important supplementary data source. A Bayesian approach (Maher, 1983) and a GLS estimator (Cascetta 1984) can be used to combine traffic counts and target demand, leading to an optimization problem as the following form:

$$\text{Min } [C' - A(D)]^T W^{-1} [C' - A(D)] + [D - D']^T Z^{-1} [D - D'] \quad (4)$$

Subject to $D \geq 0$

where W and Z denote dispersion matrices and D' is the target demand.

[Multinomial Distribution with Known Sampling Fractions]

By assuming that sample N follows a multinomial distribution with known sampling fraction α, Spiess (1987) proposed an OD demand estimation model to incorporate trip sample counts, collected from household or origin surveys. The resulting log likelihood is

$$\text{Ln } L = \text{Sum over } [N \ln(\alpha D)] + \text{constant}. \quad (5)$$

Cascetta and Nguyen (1988) gave an excellent review for estimating static OD demand matrix using traffic counts, priori information or sampling with small fractions.

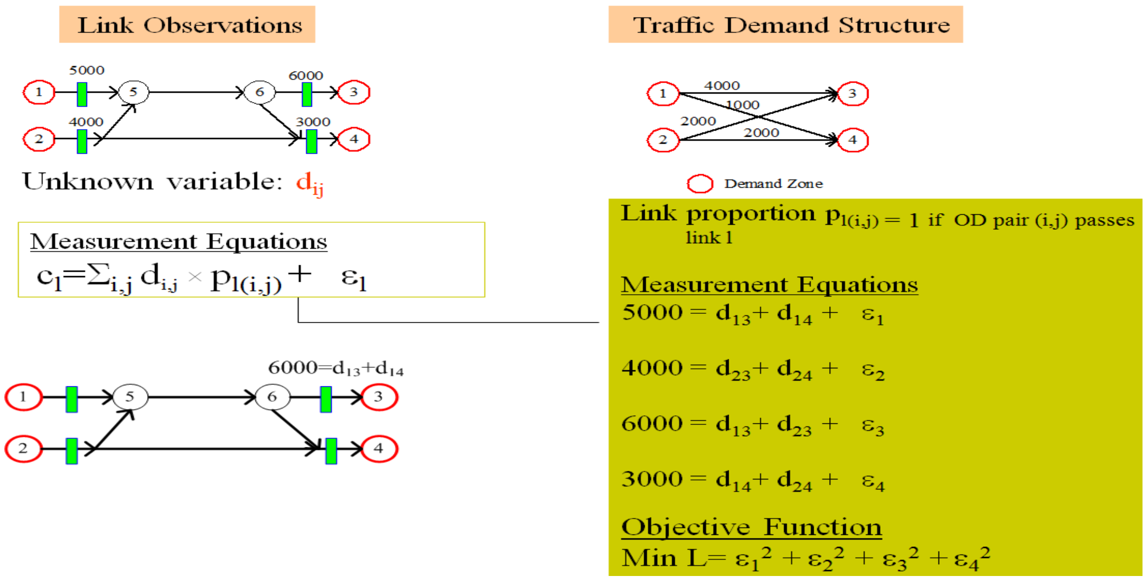
2.2 Traffic Assignment Function A and Link Flow Proportion Matrix P

The most important concept in an OD demand estimation problem is how to connect or map observations with unknown variables, in the context how to provide a mapping from OD demand d(i,j) to link counts on link l, as shown below.

$$\textbf{Link counts } \mathbf{c}_l = \mathbf{A}\ (\mathbf{d}_{(i,j)}) + \epsilon_l \quad (6)$$

where ϵ_l is the measurement noise term on link l.

Example 1: Illustration of Link Flow Proportions for OD Pair with Single Path



In the above example, there are four zones, and nodes 1, 2 are origins, and nodes 3 and 4 are destinations. Four links are equipped with loop detectors (marked as green rectangles) on all entry and exit links, with observed flow counts of 5000, 4000, 6000 and 3000, respectively. As defined below, link flow proportion matrix $P_{l,(i,j)}$ describes the proportion of flow from OD pair (i,j) passing through link *l*. For example, all traffic for OD pair from 1 to 3 has to pass through link (1->5), so the corresponding link flow proportion is 1.

By assuming the assignment function *A* gives a linear mapping between OD flow and link flow, Eq. (6) can be expressed as linear equation as

$$\mathbf{c_l=\Sigma_{i,j}\, d_{i,j} \times p_{l,(i,j)}} + \epsilon_l \qquad (7)$$

For example, link (1->5) involves two OD pairs (1 to 4 and 1 to 3), so it is easy to derive that 5000 vehicles on link (1->5) = sum of OD flow from 1 to 4 and from 1 to 3. That is, 5000 = d_{13} + d_{14} + ϵ_l, or ϵ_l = (5000 - d_{13} -d_{14})

To establish a simple OD estimation formula similar to Eq. (3), we can construct an objective function such as

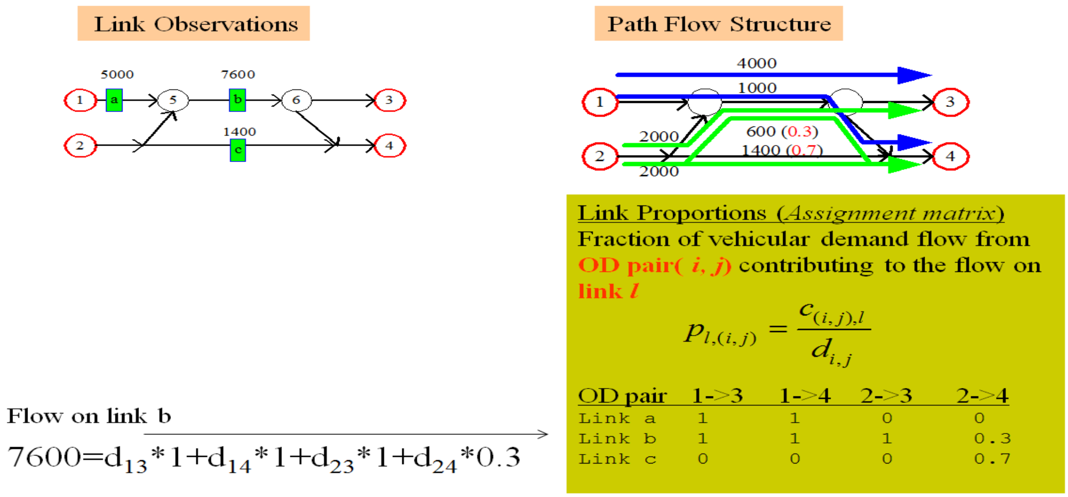
$$\min L = \sum_l [\epsilon_l]^2 \qquad (8)$$

The above example has a single path between each OD pair. When there are multiple paths on one OD pair, we can define and calculate the link proportion as

$$p_{l,(i,j)} = \frac{c_{(i,j),l}}{d_{i,j}} \qquad (9)$$

where $C_{(i,j),l}$ is the number of vehicles on link *l* coming from OD pair (i,j).

Example 2: Illustration of Link Flow Proportions for OD Pair with Multiple Paths



In Example 2, there are link sensors on links 1->5, 5->6, the frontage road along 2 to 4, denoted as links *a*, *b* and *c*. The path flow pattern is plotted in the upper right portion of the figure below. There are 2 paths from 2 to 4, 30% of flow using link b, and 70% of flow using link c. That is, link proportion from zone 1 to zone 3 on link a is 100%, from zone 2 to zone 3 on link b is 100%.

The flow on link b is sum of OD volume from 1 to 3, 1 to 4, 2 to 3 and 30% of 2 to 4, as shown as

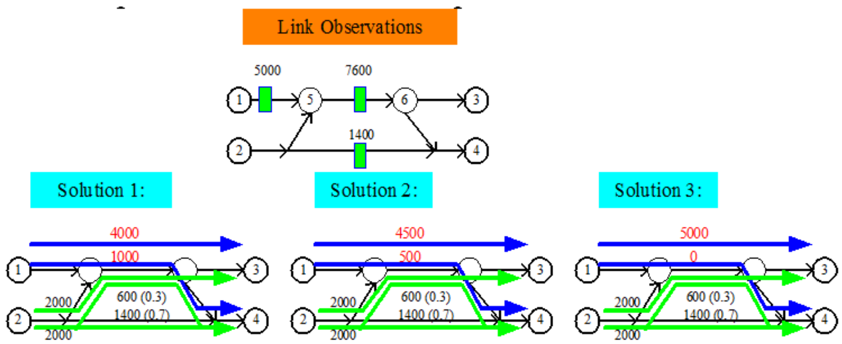
$$7600 = d_{13}*1 + d_{14}*1 + d_{23}*1 + d_{24}*0.3$$

Similarly, we can derive a mapping from OD demand d(i,j) to turning movement counts from link *l* to link *s*. **Turning Movement counts**

$$c_{l,s} = \Sigma_{i,j}\, d_{(i,j)}\, p_{l,s,(i,j)} + (l,s) \qquad (10)$$

It should be recognized that there might be multiple OD volume solutions corresponding to the same link observation pattern, as shown in Example 3 below. This is widely understood as an undetermined problem where the number of observations is less than the number of unknown independent variables (3 sensors less than 4 OD pairs in this case).

Example 3: Underdetermined Problem (# of observations ≤ # of unknown demand variables)



Problem Statement of OD Demand Estimation

Given Input:

(1) Link counts c'_l

(2) Turning movement counts $c'_{l,s}$

(3) Link flow and movement flow proportion matrices $pl_{,(i,j)}$ and $p_{l,s,(i,j)}$ (obtained from traffic assignment program)

(4) target or historical OD demand matrix $d'_{(i,j)}$

Unknown Variables:

OD demand matrix $d_{(i,j)}$

Objective Function:

When historical OD demand is available, the objective function is to minimize

(1) Deviations between **<u>estimated</u>** and **<u>observed</u>** traffic link and turning movement counts

(2) Deviations between **<u>estimated</u>** dynamic demand and **<u>target</u>** demand

$$\min L = \sum_{l}[\sum_{i,j}\left(p_{l,(i,j)} \times d_{(i,j)}\right) - c'_l]^2 + \sum_{l,s}[\sum_{i,j}\left(p_{l,s,(i,j)} \times d_{(i,j)}\right) - c'_{l,s}]^2 + \sum_{i,j}[d_{(i,j)} - d'_{(i,j)}{}^2] \qquad (11)$$

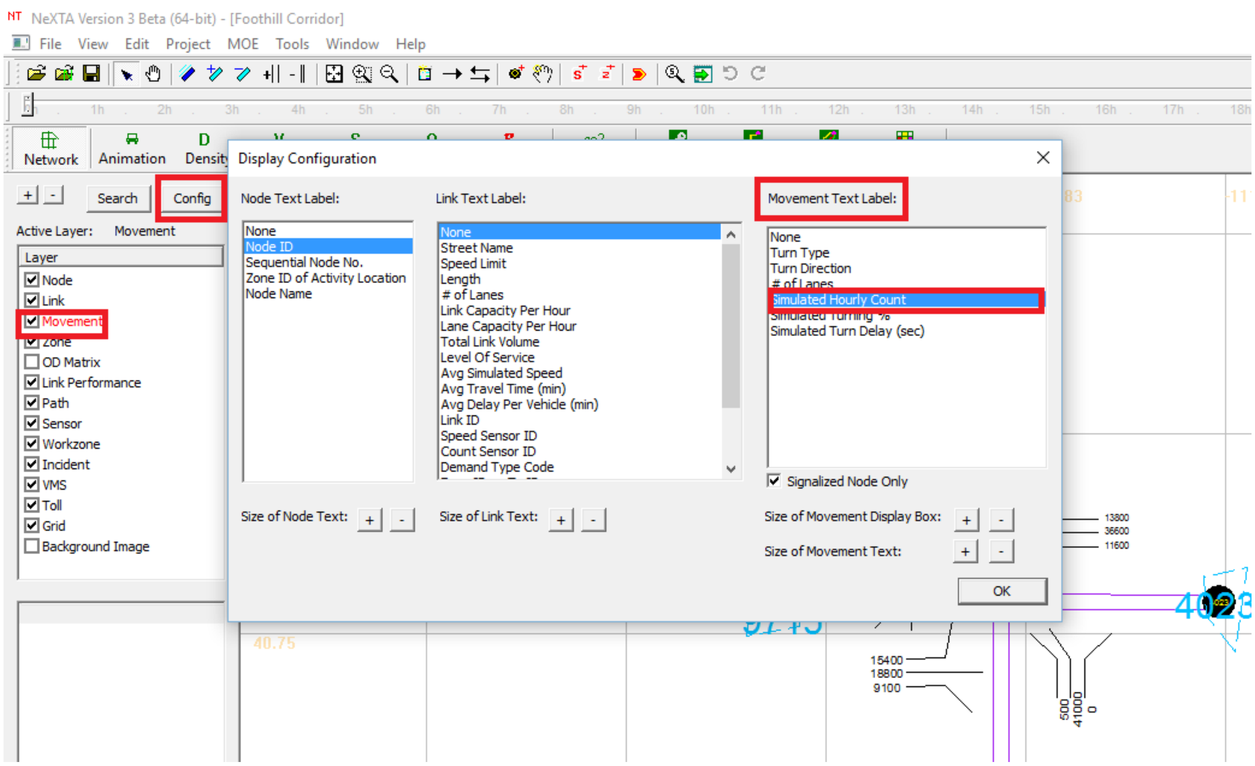
Subject to $d_{i,j} \geq 0$ ∀□, □

Task 1: Perform Static OD Demand Estimation through An Optimization Program

Step 1: Load Data Set **Foothill Corridor**

Step 2: Check turning movement counts.

Please turn on Movement GIS layer, click on Config button, and Select Observed Hourly Counts from Movement Text Label List from the Display Configuration Dialog. To change the font size, please click on "+" or "-" button.



Step 3: Locate OD demand matrix file.

The OD demand matrix is available as file input_demand.csv in the project folder. The current data are prepared by Dr. Milan Toplica Zlatkovic from the University of Utah based on his empirical information. You can also find this file directly through menu-> project ->2. demand database. Then select the first file (input_demand.csv) in the Demand File List, and click on button "Edit Selected Demand Data File in Excel".

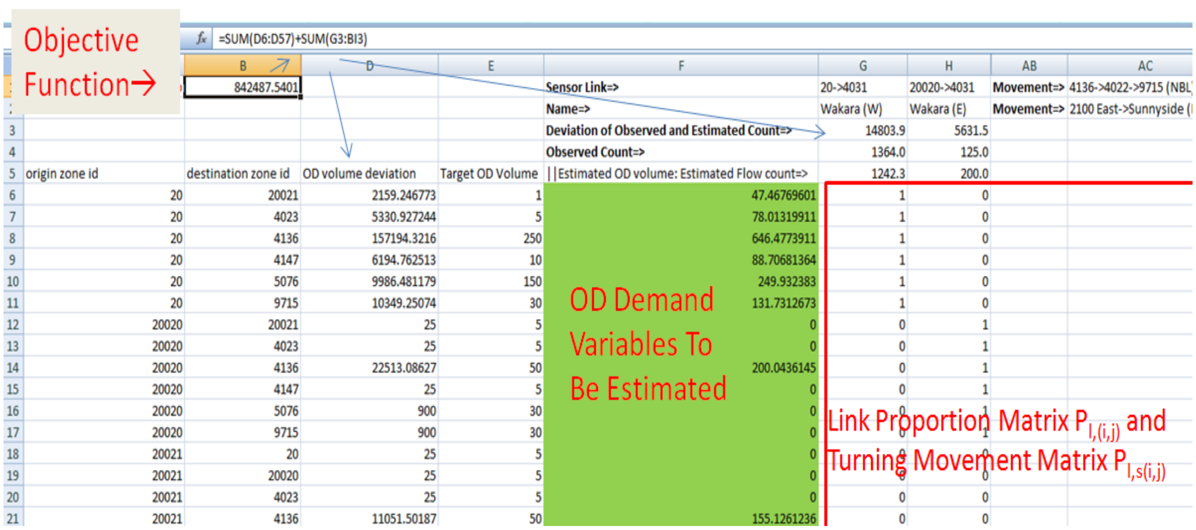
from_zone_id	to_zone_id	number_of_trips_demand_type1
5076	20021	30
5076	4147	120

5076	20020	30
5076	20	50
5076	9715	30
5076	4023	10
5076	4136	300
20021	5076	30
20021	4147	0
20021	20020	5
20021	20	5
20021	9715	30
20021	4023	5
20021	4136	50
4147	5076	150

Step 4: Generate link counts, movement counts and link/movement proportion matrix in an integrated data file.

The NEXTA/DTALite traffic assignment engine can easily produce the Link flow and movement flow proportion matrices $p_{l,(i,j)}$ and $p_{l,s,(i,j)}$.

Under GIS Layer **"Sensor"**, right click on menu "Export Link Flow Proportion Matrix to CSV File", you can find the following csv file in Excel, which has a number of built-in functions for the objective functions. Please see the online sample Google Spread file or download a copy of Exfile file with the complete optimization model here (go to Menu->File-> Download to download the original Excel File). You can go to Excel menu-> Data -> Solver to see the complete optimization model.



	A	B	D	E	F	G	H	AB	AC
1		842487.5401			Sensor Link=>	20->4031	20020->4031	Movement=>	4136->4022->9715 (NBL
2					Name=>	Wakara (W)	Wakara (E)	Movement=>	2100 East->Sunnyside (
3					Deviation of Observed and Estimated Count=>	14803.9	5631.5		
4					Observed Count=>	1364.0	125.0		
5	origin zone id	destination zone id	OD volume deviation	Target OD Volume	||Estimated OD volume: Estimated Flow count=>	1242.3	200.0		
6	20	20021	2159.246773	1	47.46769601	1	0		
7	20	4023	5330.927244	5	78.01319911	1	0		
8	20	4136	157194.3216	250	646.4773911	1	0		
9	20	4147	6194.762513	10	88.70681364	1	0		
10	20	5076	9986.481179	150	249.932383	1	0		
11	20	9715	10349.25074	30	131.7312673	1	0		
12	20020	20021	25	5	0	0	1		
13	20020	4023	25	5	0	0	1		
14	20020	4136	22513.08627	50	200.0436145	0	1		
15	20020	4147	25	5	0	0	1		
16	20020	5076	900	30	0	0	1		
17	20020	9715	900	30	0	0	1		
18	20021	20	25	5	0	0	0		
19	20021	20020	25	5	0	0	0		
20	20021	4023	25	5	0	0	0		
21	20021	4136	11051.50187	50	155.1261236	0	0		

Remarks:

1. At rows 1 and 2, all link and movement counts (c'_l and $c'_{l,s}$) are listed column by column with the corresponding street names and directions.

2. Observed link and movement counts are listed in row 4. Row 5 calculates the estimated counts, by using an Excel built-in function SUMPRODUCT() to implement the measurement function

$\mathbf{c_l = \Sigma_{i,j}\, d_{i,j} \times p_{l(i,j)}}$ and $\mathbf{c_{(l,s)} = \Sigma_{i,j}\, d_{i,j} \times p_{l,s\,(i,j)}}$

3. At columns A to E, All OD demand pairs are listed row by row, with prior data as “target demand matrix” $\mathbf{d_{i,j}'}$ by default.

4. Unknown demand variables $\mathbf{d_{i,j}}$ are listed in column F, and column D corresponds to equations

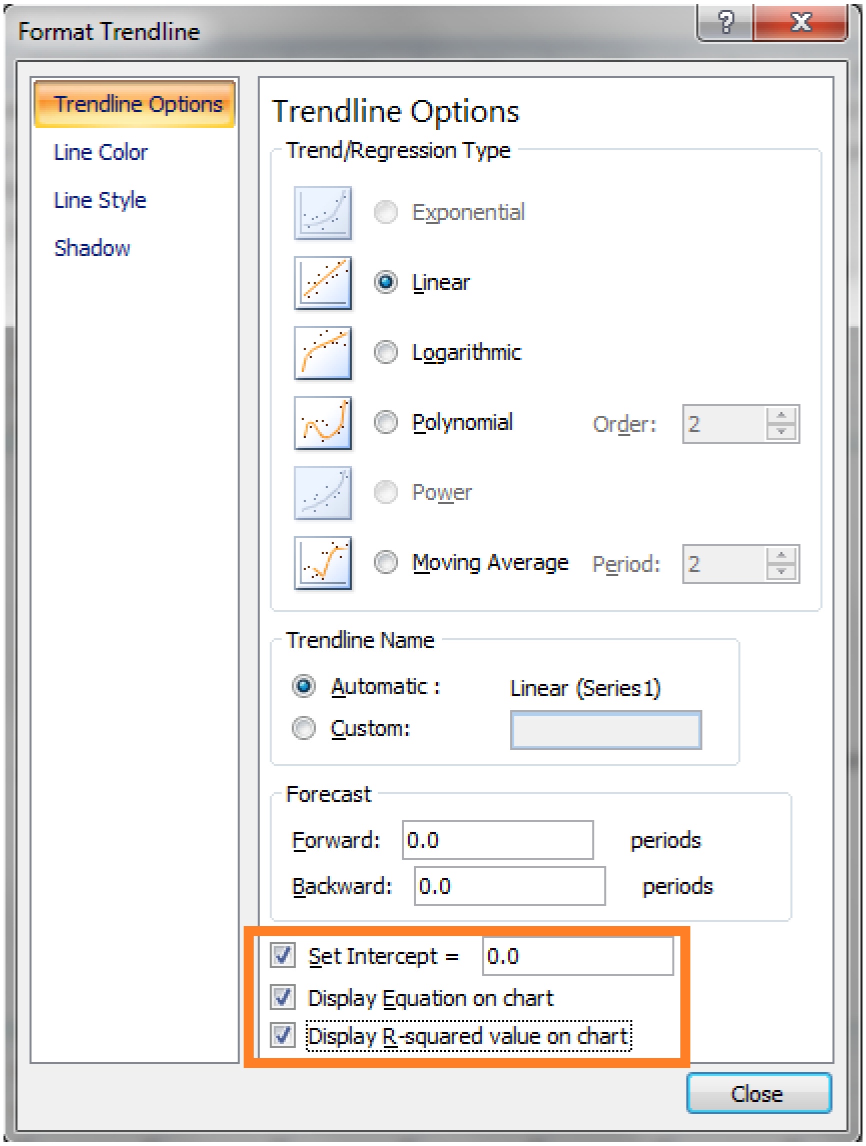
$$\sum_{i,j}[d_{(i,j)} - d'_{(i,j)}{}^2]$$

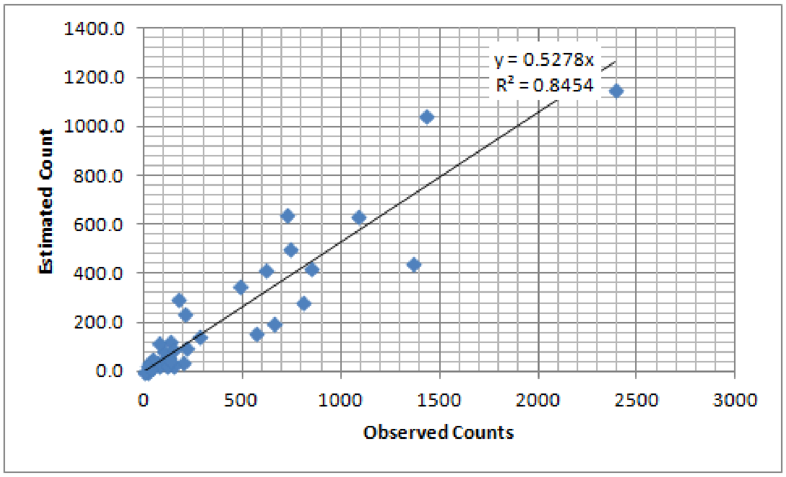
5. The objective function (11) is coded in Cell B2.

Step 5: Compare observed link/movement counts and “simulated” link counts from the given OD demand matrix.

You can first measure the existing deviation between observed counts and estimated counts (based on Milan’s prior OD demand table).

Select data block in rows 4 and 5, and construct a scatter plot with the a trendline. Please format the trend line as the following.

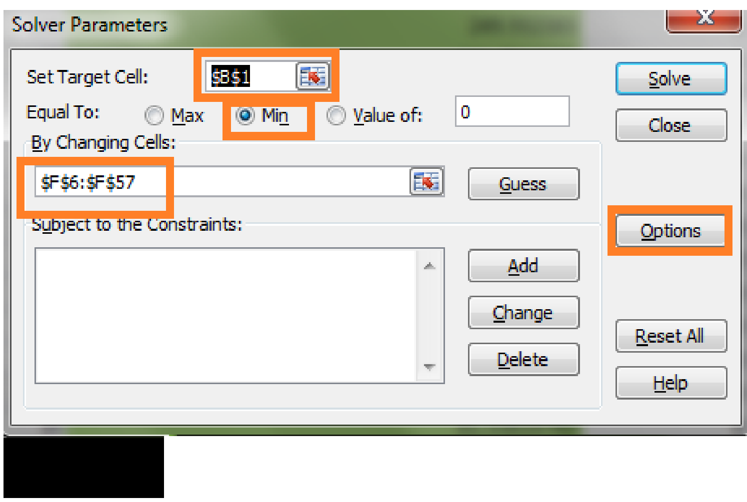

Format Trendline
Trendline Options
Line Color
Line Style
Shadow
Trendline Options
Trend/Regression Type
Exponential
Linear
Logarithmic
Polynomial
Order: 2
Power
Moving Average
Period: 2
Trendline Name
Automatic : Linear (Series1)
Custom:
Forecast
Forward: 0.0 periods
Backward: 0.0 periods
Set Intercept = 0.0
Display Equation on chart
Display R-squared value on chart
Close

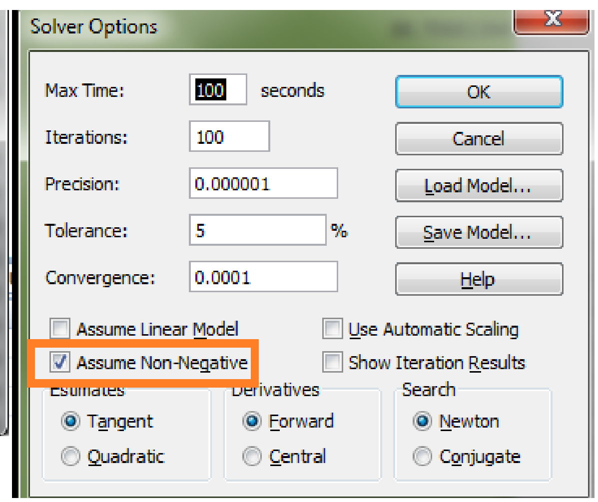


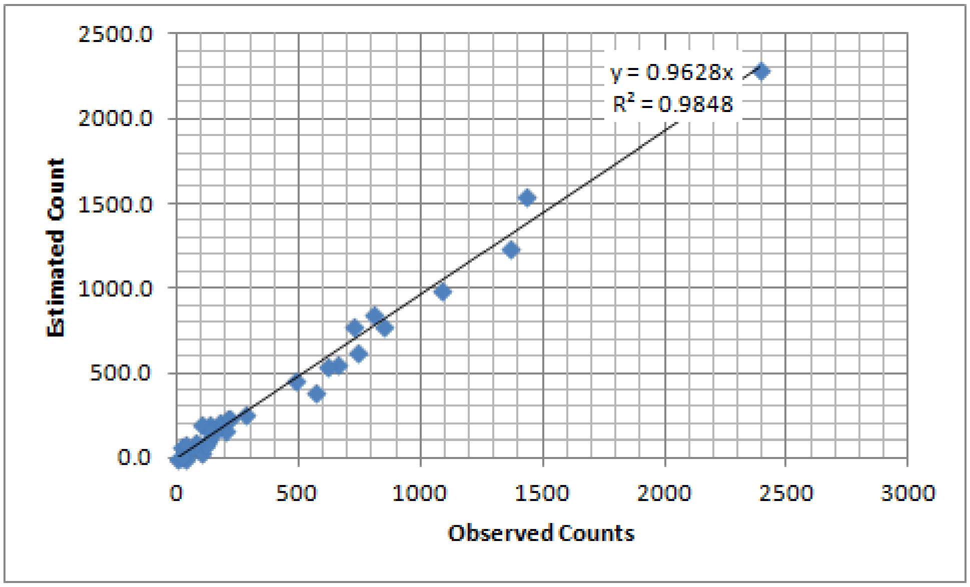
Y=0.5278 x in the above chart indicates that the current OD demand significantly underestimates the ground truth traffic.

Step 6: Construct optimization program for OD demand estimation

The construct an optimization model, please define target cell as B1, use Min, Select the green block highlighted above as the "changing cells". Click on the Options button, and check Assume Non-negative.





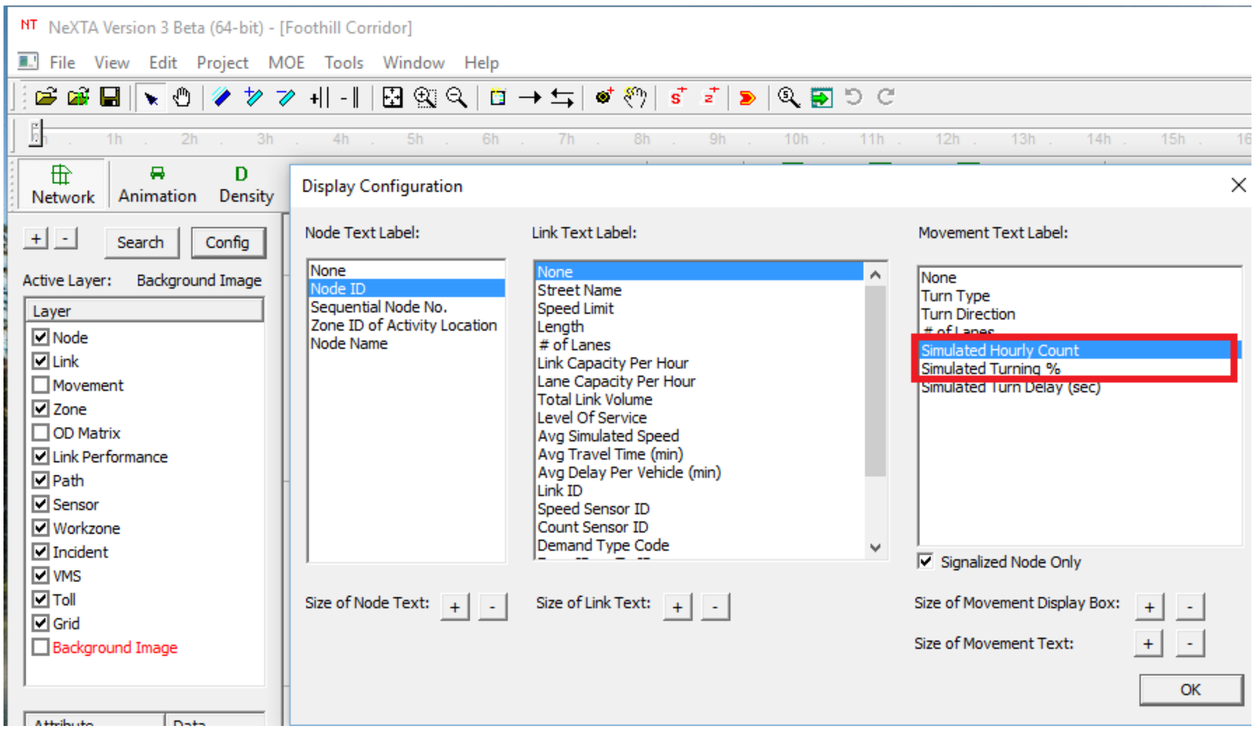


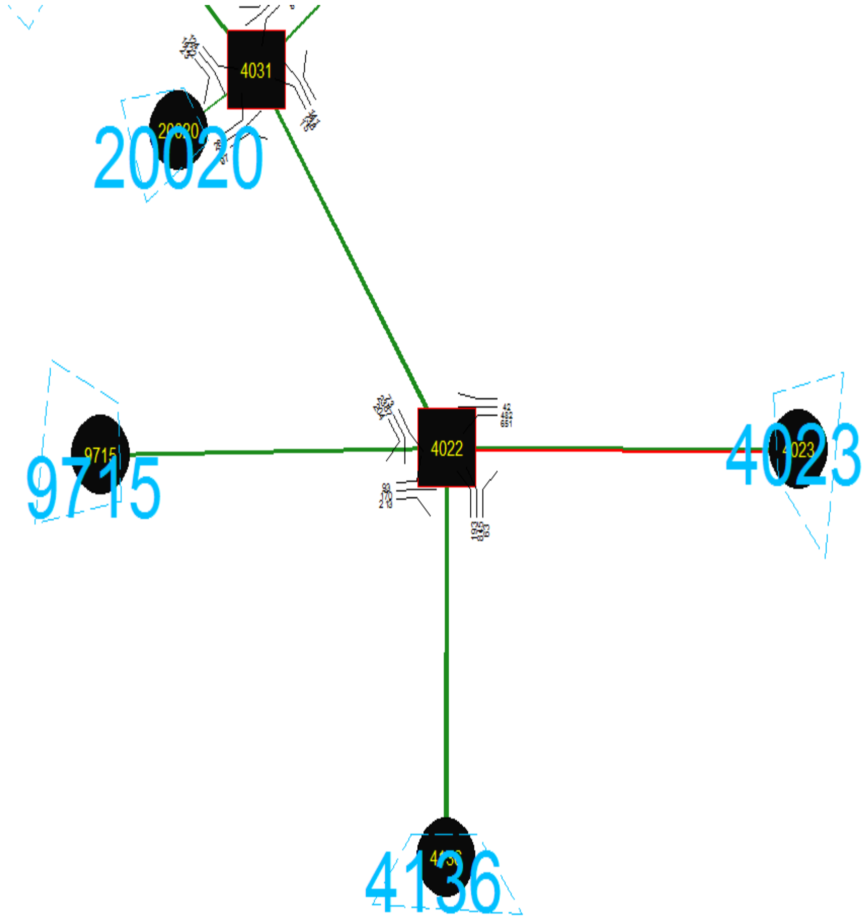
After the OD demand estimation step, the slope of 0.9628 in Y=0.9628 x in the above chart shows a great improvement in the overall fitness, while 1 is the ideal ratio. The R^2 also means very good fit to link and turning movement counts.

Task 2: Estimate OD Demand Matrix Based on Entry Link Volume and Turning Proportions

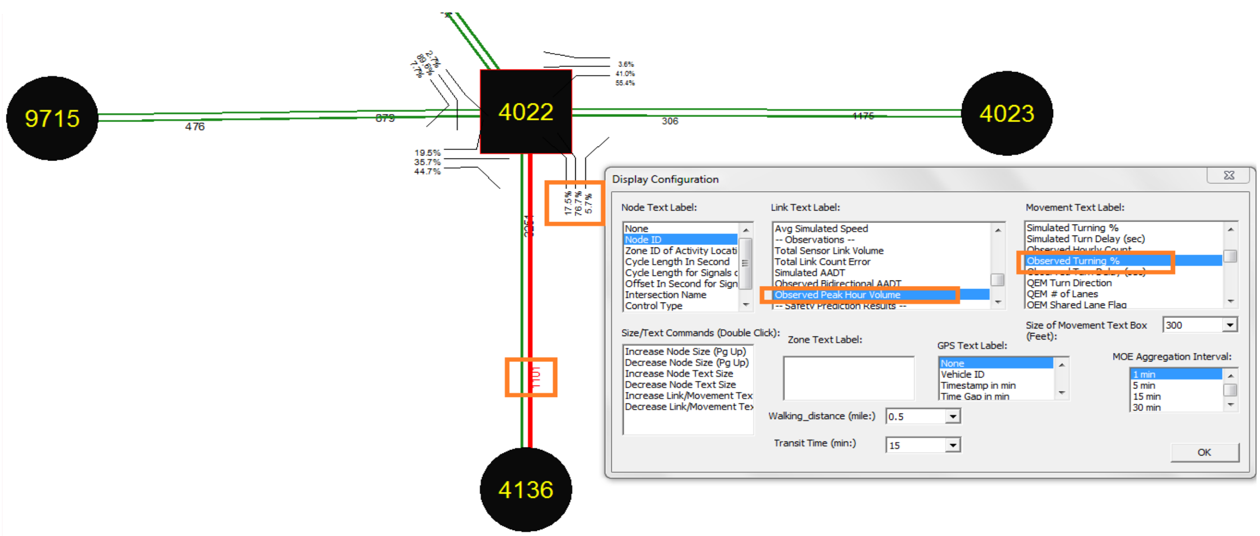
This task aims to help students understand how a small-scale OD demand matrix can be manually prepared based on entry link flow volume and turning movement percentage. This process has been widely used by many practitioners but it has certain limitations such as only suitable for small networks and unable to handle possible counting inconsistency between different locations.

Step 1: Use NEXTA to show the *observed hourly turning counts* and *turning percentages* through the Display Configuration dialog.





Step 2: Show the *<u>observed Peak Hour Link Volume</u>* through the Display Configuration dialog.

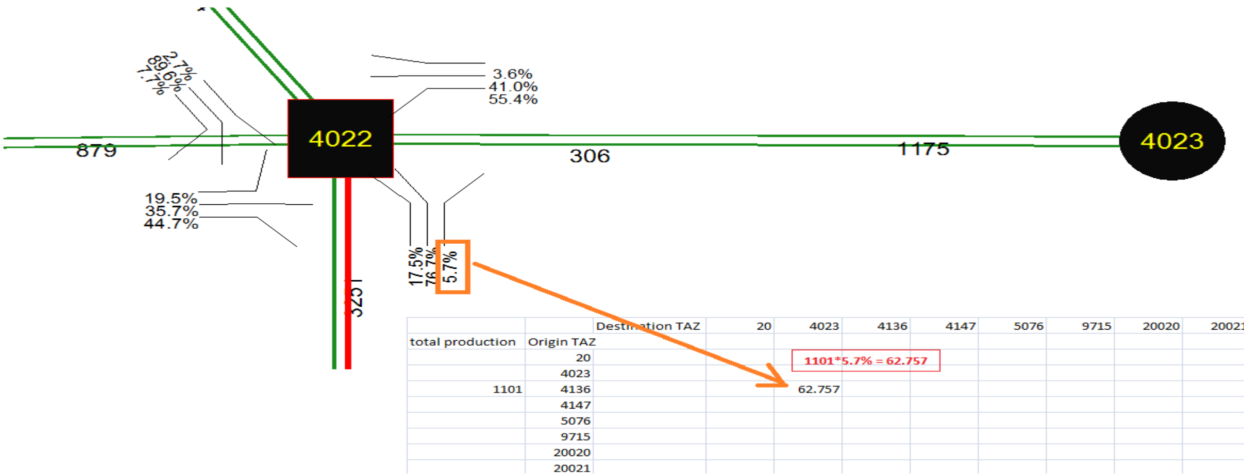


[AA1]

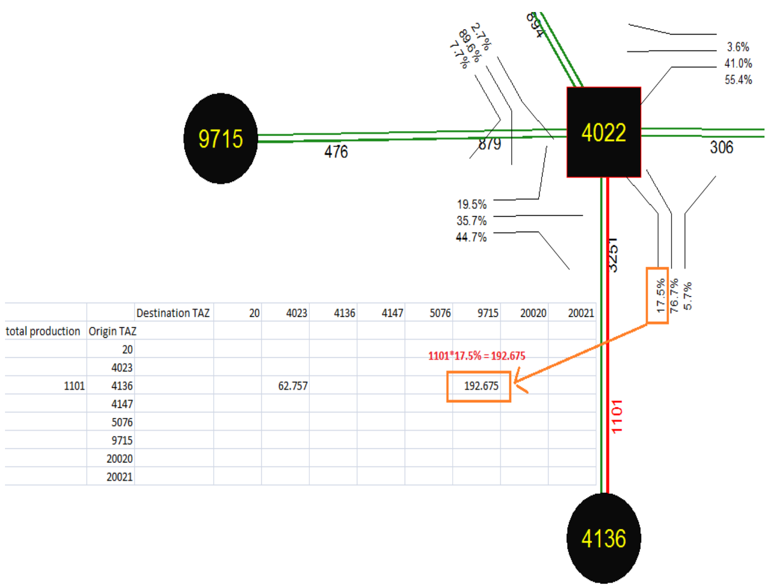
Step 3: Construct an OD demand matrix as shown below, where zone 4136 has a total outflow of 1101 vehicles from the above figure. An online spreadsheet can be downloaded from here.

		Destination TAZ	20	4023	4136	4147	5076	9715	20020	20021
total production	Origin TAZ									
	20									
	4023									
1101	4136									
	4147									
	5076									
	9715									
	20020									
	20021									

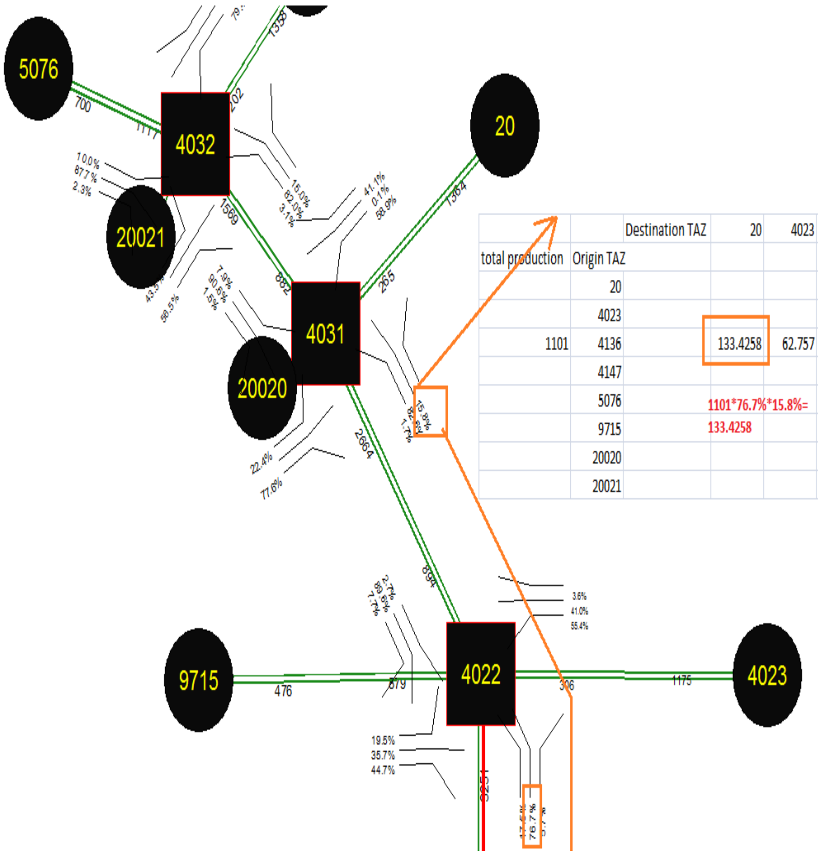
Step 4: Estimate OD demand volume from zone 4136 to 4023 by using link count (1101) times turning percentage (5.7%) = 62.757.



Step 5: Estimate OD demand volume from zone 4136 to 9715 by using link count (1101) times turning percentage (17.5%) = 192.675.



Step 6: Estimate OD demand volume from zone 4136 to 20 (going through path 4136, 4022, 4031 and 20) by using link count (1101) times first through percentage (17.5%) and the second right turn percentage (15.8%) = 133.4248.

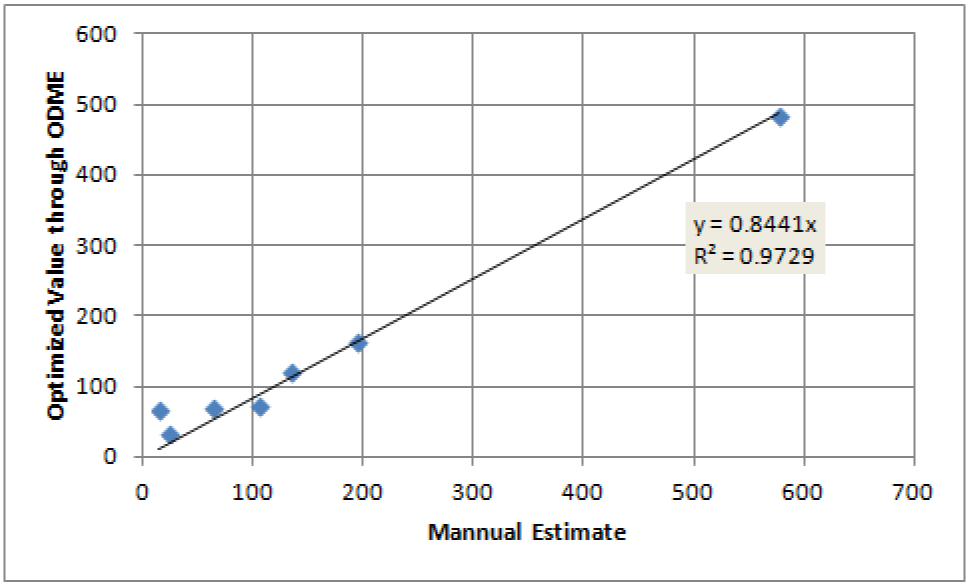


Step 7: By assuming same turning percentages for all OD pairs, one can estimate the OD flow counts from zone 4036 to all other zones.

		Destination TAZ	20	4023	4136	4147	5076	9715	20020	20021
total production	Origin TAZ									
	20									
	4023									
1101	4136		134	63		105	577	193	14	22
	4147									
	5076									
	9715									
	20020									
	20021									

Step 8: Compare your manually estimated OD volume with the "optimized" OD volume for zone 4136. You can find the manual estimates are quite consistent with the optimized values from OD Matrix Estimation program.

Destination TAZ	20	4023	4136	4147	5076	9715	20020	20021	20020	20021
Mannul Estimate	134	63		105	577	193	14	22		
Optimized Value	122	70		72	483	165	67	35		



Step 9: Please manually develop the OD demand matrix using the above procedure for all the other zones. Submit your solution. Please also offer your observations, such as 1. Inconsistent counts between intersections; 2. The overall link count and movement count estimation quality for manual and optimal estimates.

Reference:

Van Zuylen and Willumsen(1980，the most likely trip matrix estimated from traffic counts)

Maher (1983, inference on trip matrices from observations on link volumes: a Bayesian statistical approach)

Cascetta (1984, estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator)

Spiess (1987, a maximum likelihood model for estimating origin-destination matrices)

Cascetta and Nguyen (1988, a unified framework for estimating or updating origin/destination matrices from traffic counts)

Recommendations:

Please send your comments to xzhou99@gmail.com, as we would like to improve this working document to benefit other users/students! (You can also directly add comments in this online learning document.) Thank you!