Nature and Causes of Urban Traffic Congestion –

A Data Analytic Approach

Andy Chow

Systems Engineering and Engineering Management

City University of Hong Kong

Data analytics

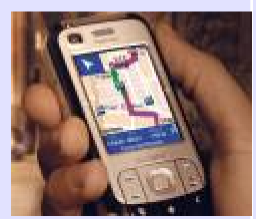
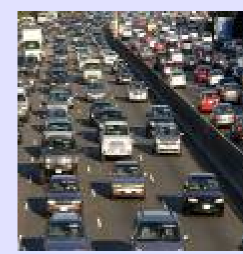
- Knowledge of urban traffic characteristics is prerequisite for effective congestion management

- Most research on traffic analysis has been focusing on motorways / freeways
  - lack of required urban traffic data
  - complexity of the urban network

- Increased availability of urban traffic data (e.g. loop detectors, GPS, vehicle identification, etc) provides new research opportunities

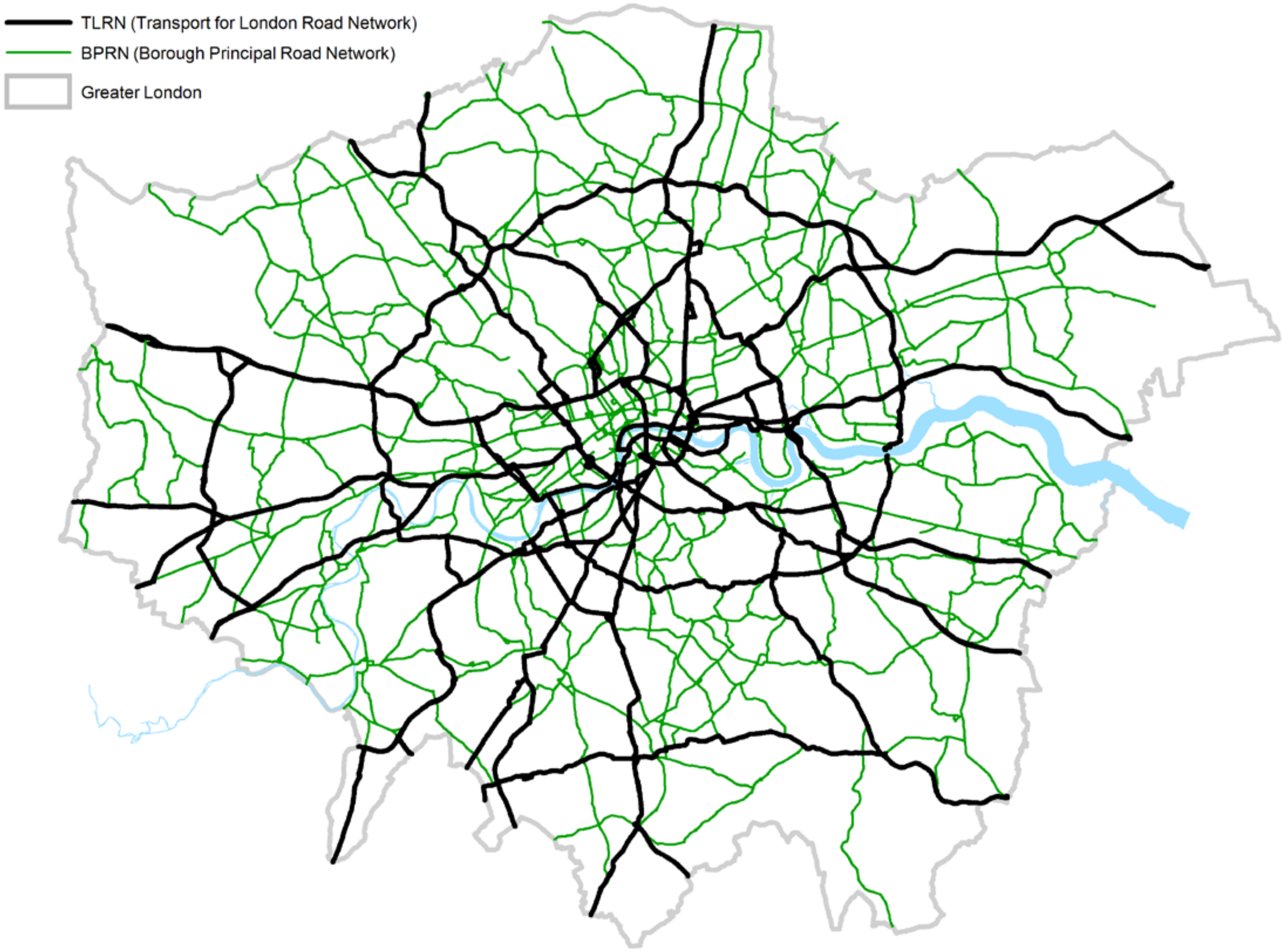
Data analytics

- Fixed sensors (e.g. loop detectors, traffic counters)
- GPS devices (e.g. iBus, Addison Lee Taxi data)
- Automatic Vehicle Identification (e.g. ANPR)

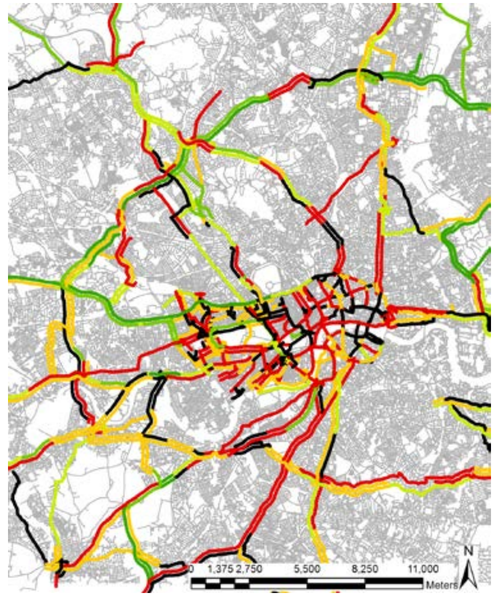
Fixed	GPS (mobile)	AVI (Spatial)
 	 	 

London Congestion Analysis Project (LCAP)

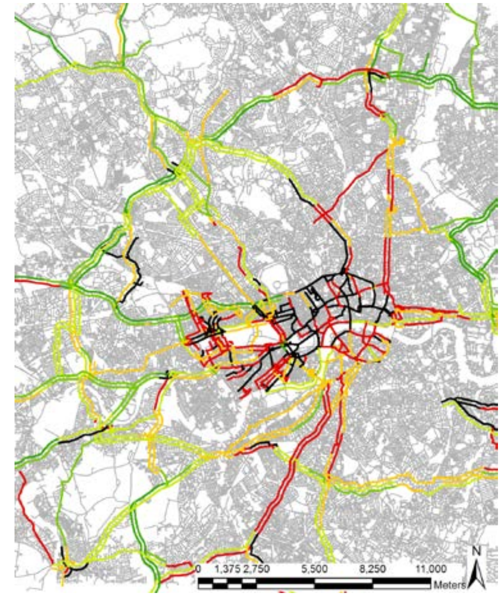
- A total of 500 cameras in London for enforcing different schemes
- Number plates recorded, with associated time stamps
- Derive the vehicle travel times by matching number plates by using **Automatic Number Plate Recognition (ANPR) technique**



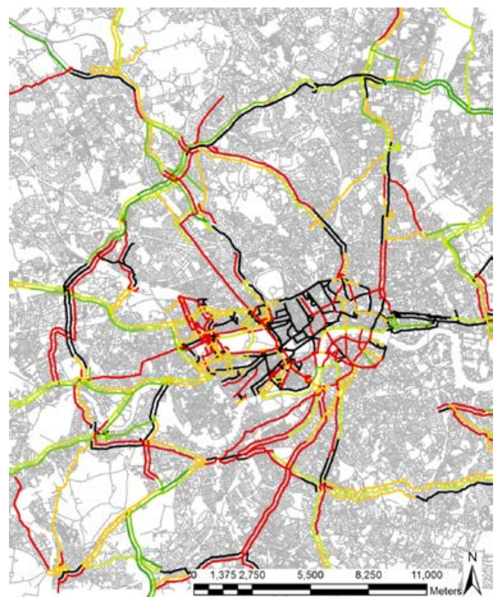
London Road Network



08:00



10:00

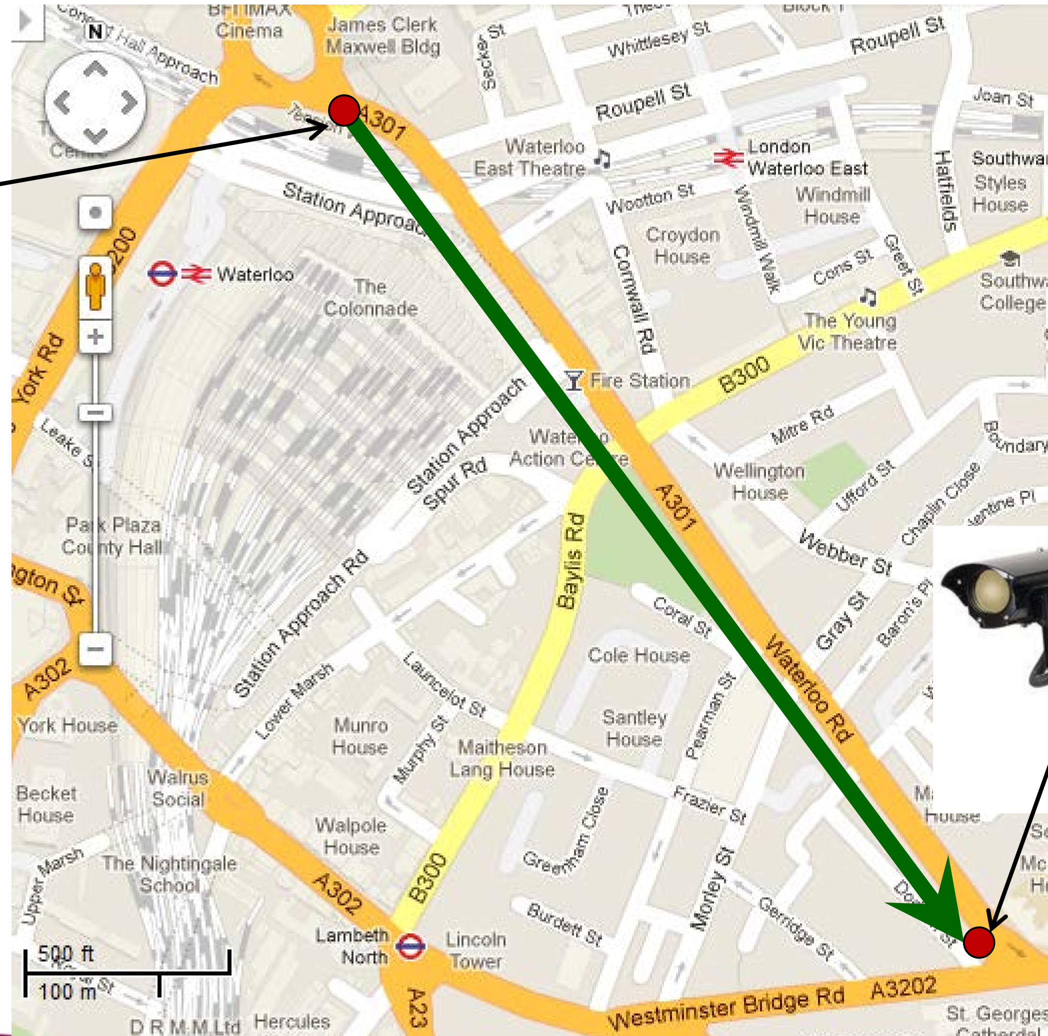


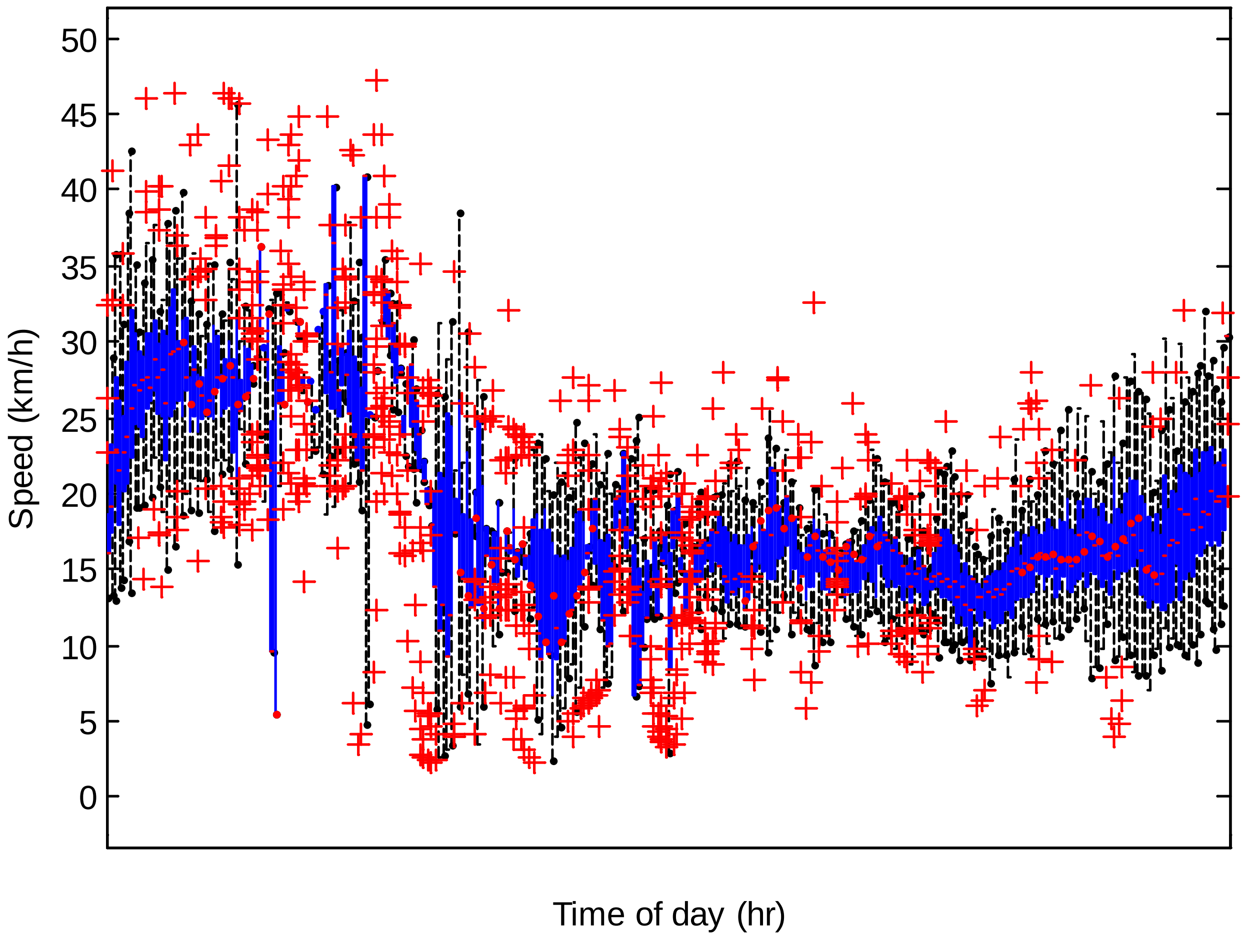
17:00

5 minute aggregated interval travel Time (mins/km)

- 0.324000 - 1.220000
- 1.220001 - 1.711000
- 1.711001 - 2.335000
- 2.335001 - 3.133000
- 3.133001 - 4.395000
- 4.395001 - 35.533000

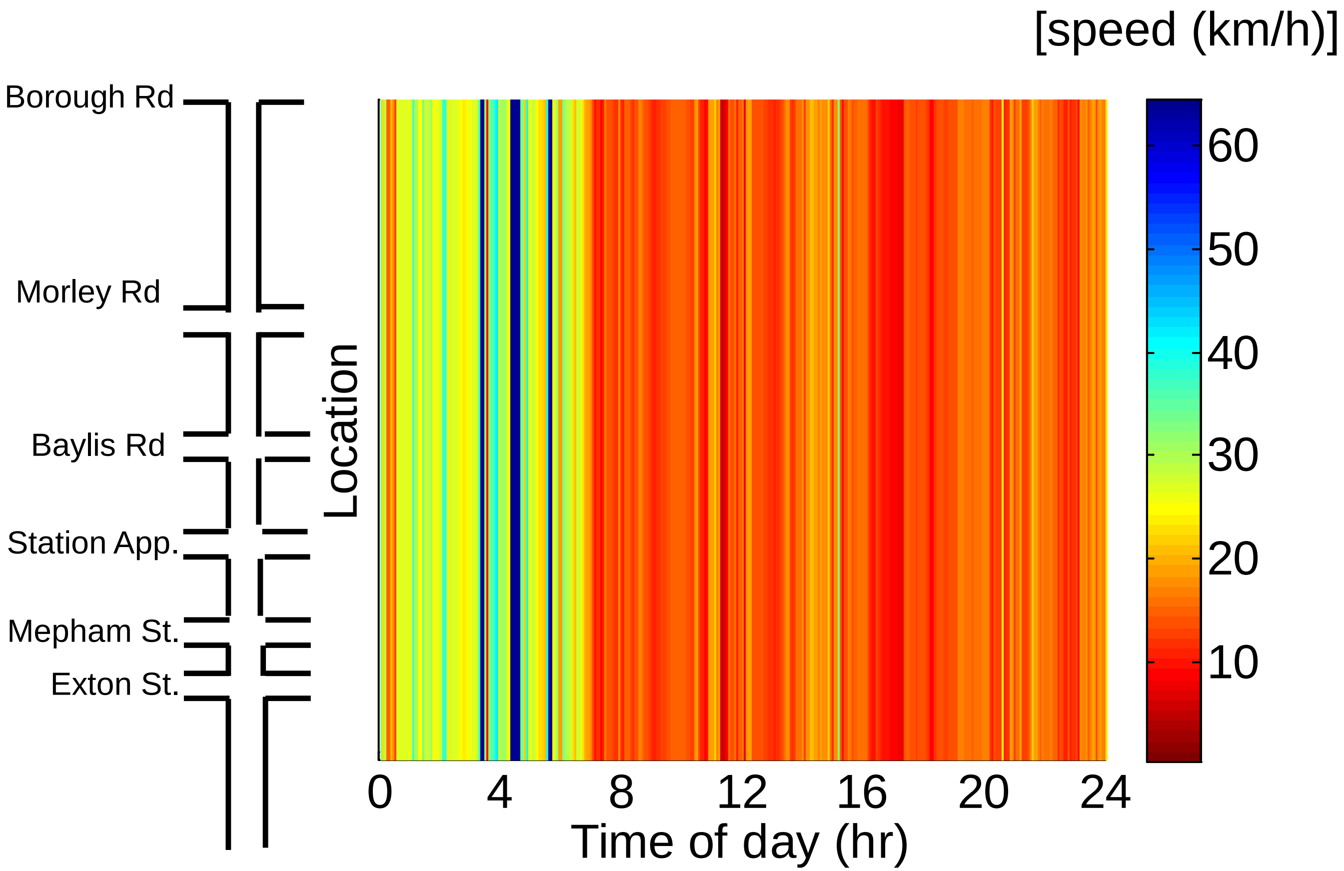
Waterloo Rd (Southbound), Central London




50
45
40
35
30
25
20
15
10
5
0
Speed (km/h)
Time of day (hr)

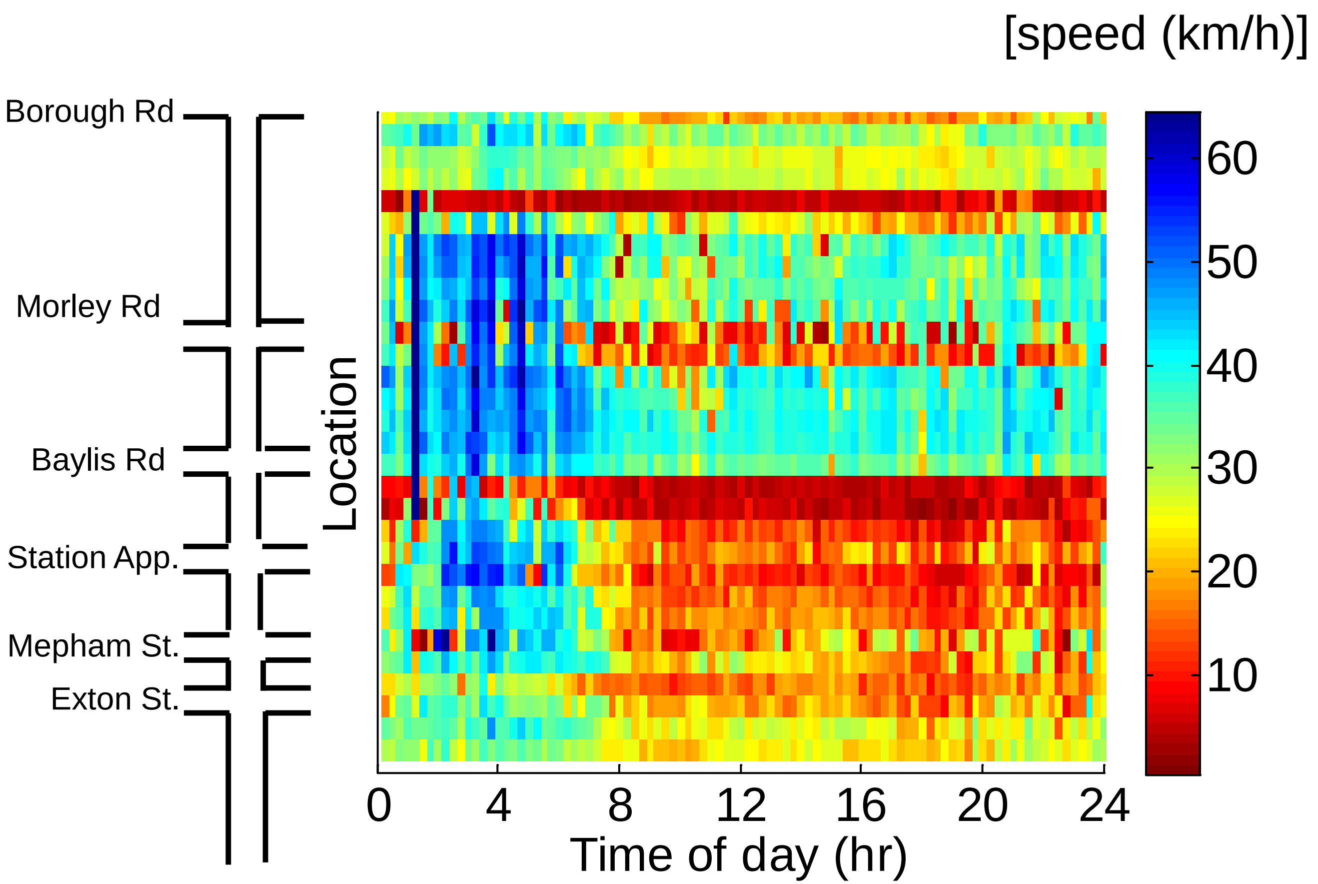
ANPR data

- (derived from journey times over a day)



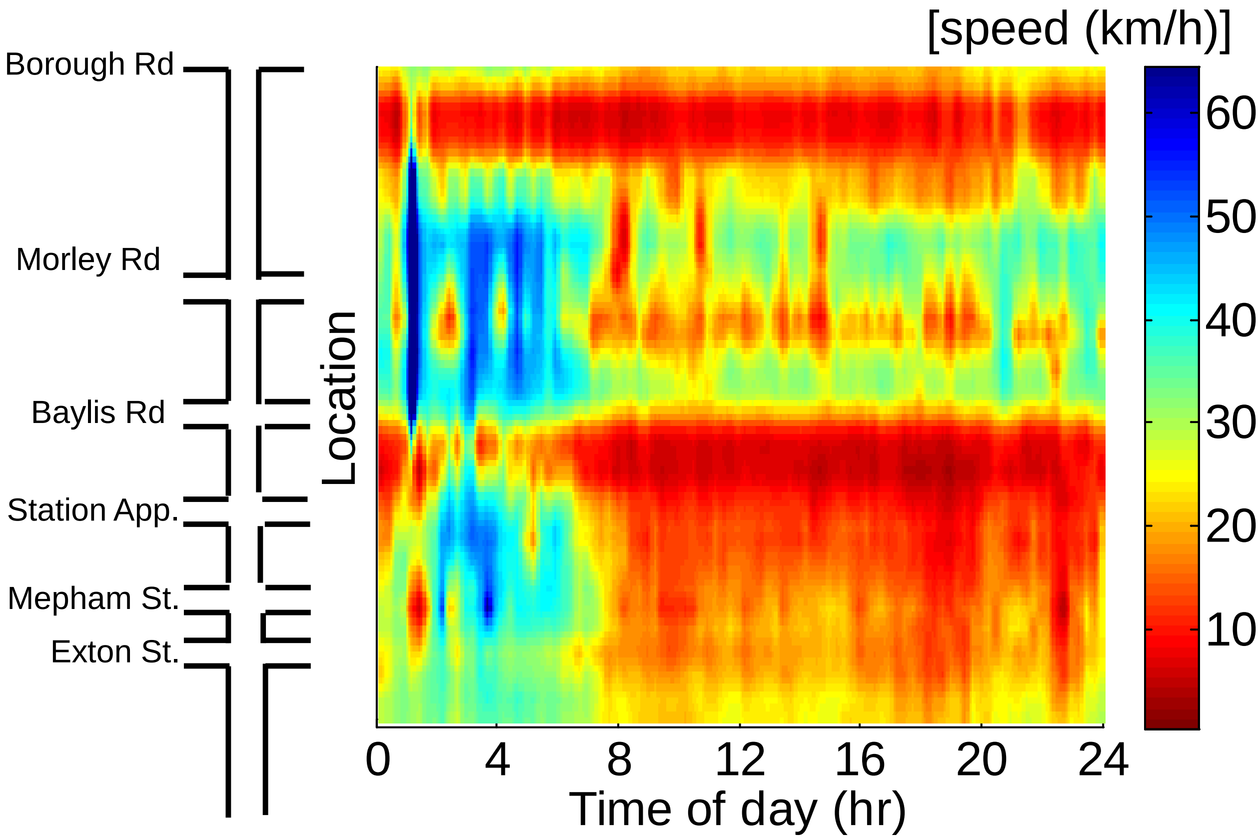
GPS probe data (trafficmaster, Addison Lee)

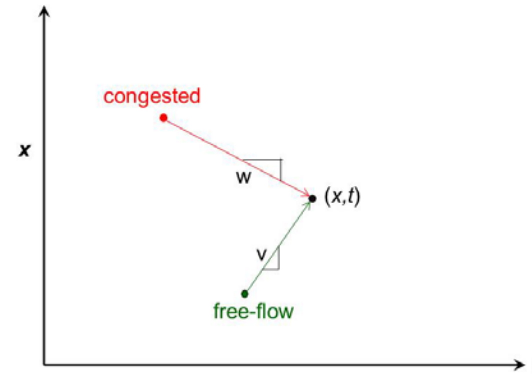
- (collected over a month period…)



Data fusion (state reconstruction)

(ANPR + GPS)





Adaptive Smoothing Method (ASM) – (Treiber-Helbing)

- 'Anisotropic' formulae:

$$u_{free}(x,t) = \frac{1}{N(x,t)} \sum_{i=1}^{M(x,t)} \phi_i\left(x - x_i, t - t_i - \frac{x-x_i}{v}\right) u_i$$

$$u_{cong}(x,t) = \frac{1}{N(x,t)} \sum_{i=1}^{M(x,t)} \phi_i\left(x - x_i, t - t_i - \frac{x-x_i}{w}\right) u_i$$

$$u(x,t) = \gamma(x,t) u_{cong}(x,t) + [1-\gamma(x,t)] u_{free}(x,t)$$

Data integration

- The smoothed data can then be integrated as

$$\tilde{u}(x,t) = \frac{1}{E(x,t)} \sum_{k=1}^{K(x,t)} \varepsilon_k(x,t) u_k(x,t)$$

- Weighting of data depends on various factors: sample size, data variance, and quality of the source data (e.g. level of imputation)

Reference:

Chow, A.H.F., et al. (2013) Analysis of adaptive data fusion algorithm for urban network application. Proceedings of the 92nd Annual meeting of the Transportation Research Board (Paper #: 13-4189), January 13-17. Washington, DC, USA.

CityU

Analysis – causes of congestion

- To investigate the nature and causes of urban congestion with different sources of data
- Performance measures
  - Journey time (measured from ANPR)
- Level of demand
  - Oyster counts
- Non-recurrent factors
  - Incident and weather records.

STANDARD
Spatio-Temporal Analysis of Network Data and Route Dynamics

Non-recurrent factors (LSTCC)

- The London Streets Traffic Control Centre (LSTCC) records the following incidents:
  - Serious Accidents (Event-Hrs)
  - Moderate Accidents: (Event-Hrs)
  - Breakdowns: (Event-Hrs)
  - Obstruction: (Event-Hrs) due to obstructions other than broken down vehicles. E.g. fallen tree(s) or signal failures;
  - Police checks: (Event-Hrs)
  - Special events: (Event-Hrs) such as filming, car race, carnival, demonstrations, football matches, etc;
  - Roadwork (Event-Hrs)

Regression analysis

- Congestion – represented by sum of journey times in *Central London* on each day *d:*

$$Y_{total}(d)$$

- Expressed as a function of a list of explanatory variables on each day *d*

$$Y_{total}(d) = \beta_0 + \sum_i \beta_i X_i(d) + \varepsilon(d)$$

$X_{oyster}(d)$	Total number of bus oyster counts recorded on day d
$X_{acc_{m}od}(d)$	Total duration (Event-Hrs) of moderate accidents on day d
$X_{acc_{s}ns}(d)$	Total duration (Event-Hrs) of serious accidents on day d
$X_{brk}(d)$	Total duration (Event-Hrs) affected by vehicular breakdowns on day d
$X_{obs}(d)$	Total duration (Event-Hrs) affected by road obstruction on day d
$X_{roadwork}(d)$	Total duration (Event-Hrs) of road work on day d
$X_{event}(d)$	Total duration (Event-Hrs) of special events on day d
$X_{police}(d)$	Total duration (Event-Hrs) of police security checks on day d
$X_{rain}(d)$	Total precipitation (mm) measured on day d
$X_{strike}(d)$	An 0-1 indicator which equals 1 if tube strike on day d; 0 otherwise
$X_{snow}(d)$	An 0-1 indicator which equals 1 if snow on day d; 0 otherwise

Regression analysis

- Congestion – represented by sum of journey times in *Central London* on each day *d:*

$$Y_{total}(d)$$

- Expressed as a function of a list of explanatory variables on each day *d*

$$Y_{total}(d) = \beta_0 + \sum_i \beta_i X_i(d) + \varepsilon(d)$$

- The parameters β_i reflect the sensitivity and significance of each factor on 'congestion'

Scenario	Factors	Coeff. (β)	Std. Error	*t*- value	*p*- value		R^2
AM	(Intercept)	211343.4	6338.1	33.345	0.000	***	0.61
	Bus Oyster counts	0.0628	0.0020	31.528	0.000	***	
	Moderate Accidents	3102.3	695.2	4.463	0.000	***	
	Serious Accidents	6558.2	1313.8	4.992	0.000	***	
	Vehicular breakdowns	4402.3	967.7	4.549	0.000	***	
	Obstruction	149.2	317.0	0.471	0.633		
	Roadwork	161.4	74.2	2.176	0.003	**	
	Special events	499.8	165.6	3.017	0.001	***	
	Police checks	-955.2	1381.8	-0.691	0.492		
	Strike	99972.6	23429.8	4.267	0.000	***	
	Snow	-11232.4	8126.5	-1.382	0.169		
	Precipitation	1502.3	5142.9	0.292	0.789		
PM	(Intercept)	227837.3	8124.9	28.042	0.000	***	0.62
	Bus Oyster counts	0.0759	0.0026	29.216	0.000	***	
	Moderate Accidents	2742.4	695.7	3.942	0.000	**	
	Serious Accidents	6050.1	1015.2	5.960	0.000	**	
	Vehicular breakdowns	3393.9	1085.4	3.127	0.000	***	
	Obstruction	161.8	329.9	0.490	0.629		
	Roadwork	132.0	56.5	2.339	0.018	*	
	Special events	848.0	147.0	5.770	0.000	***	
	Police checks	1255.5	1162.1	1.080	0.338		
	Strike	67268.1	18276.4	3.681	0.002	**	
	Snow	-17042.3	6884.7	-2.475	0.032	*	
	Precipitation	-112.1	1705.2	-0.066	0.945		

Regression analysis

- Expressed as a function of a list of explanatory variables on each day *d*

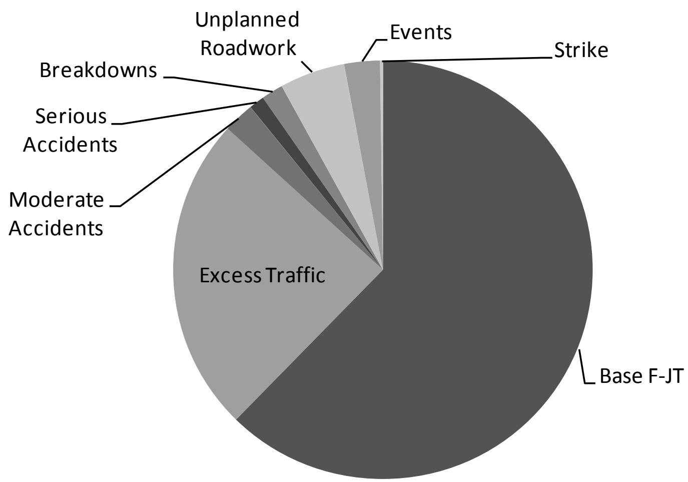
$$Y_{total}(d) = \beta_0 + \sum_i \beta_i X_i(d) + \varepsilon(d)$$

- Hence

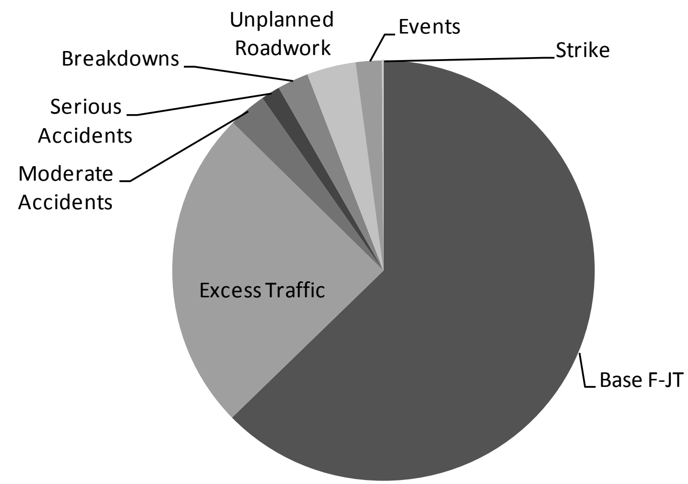
$$\overline{Y}_{total} = \beta_0 + \sum_i \beta_i \overline{X}_i$$

gives the estimated components of congestion due to different factors

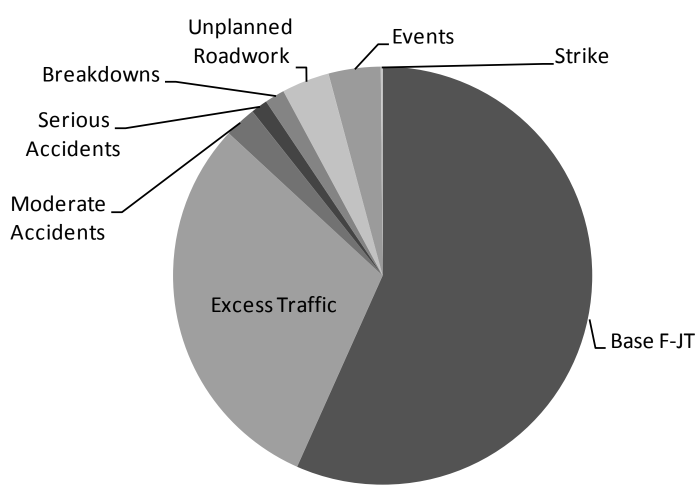
Components of congestion



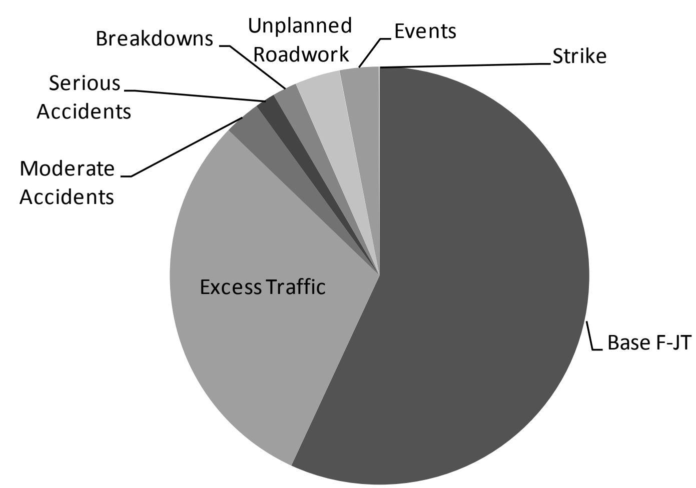
AM



AM



PM
2009



PM
2010

Concluding remarks

- The regression analysis identifies the significant factors that will have an impact on the journey times (congestion)
- It is found that 25%-30% of the observed congestion is due to travel demand, while 15% of the observed journey times is due to the non-recurrent factors (e.g. accidents, roadwork, special events, strikes)
- This provides insights on how we should make investment on the transport infrastructure.