SATURN UGM 2014

New data sources & data fusion

Andrew Merrall

14th November 2014

Plan Design Enable

Objectives, background

- Assess efficacy of new data sources in enhancing a model's 'prior matrix'
- RSI too difficult
- New [Big] data – mobile, GPS, wifi, Bluetooth

Overall process

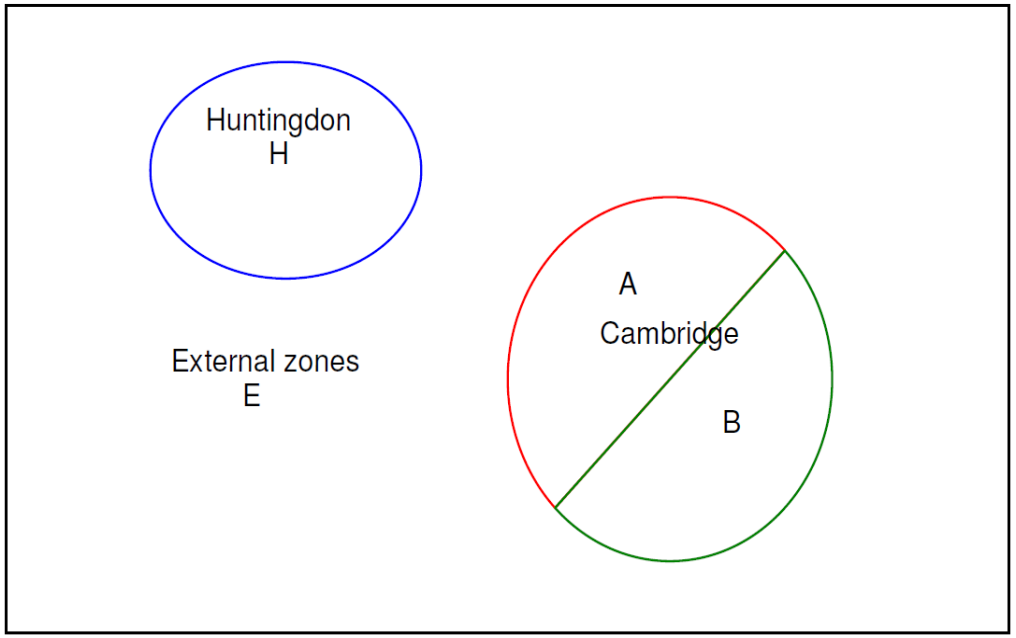
- Old model & new data
- Data processing
- Testing

Model test bed

- Partial & synthetic matrices in line with good practice
- Combined matrices allow for statistical reliability
- Screenlines for matrix validation
- ME has impacted on prior
- ME applied at mini-screenlines or cordons
- Acceptable to new data source

CSRM

- Multi mode model, 325 zones
- Base year 2006, 10 user class, 2 vehicle types, 3 purposes, 3 incomes



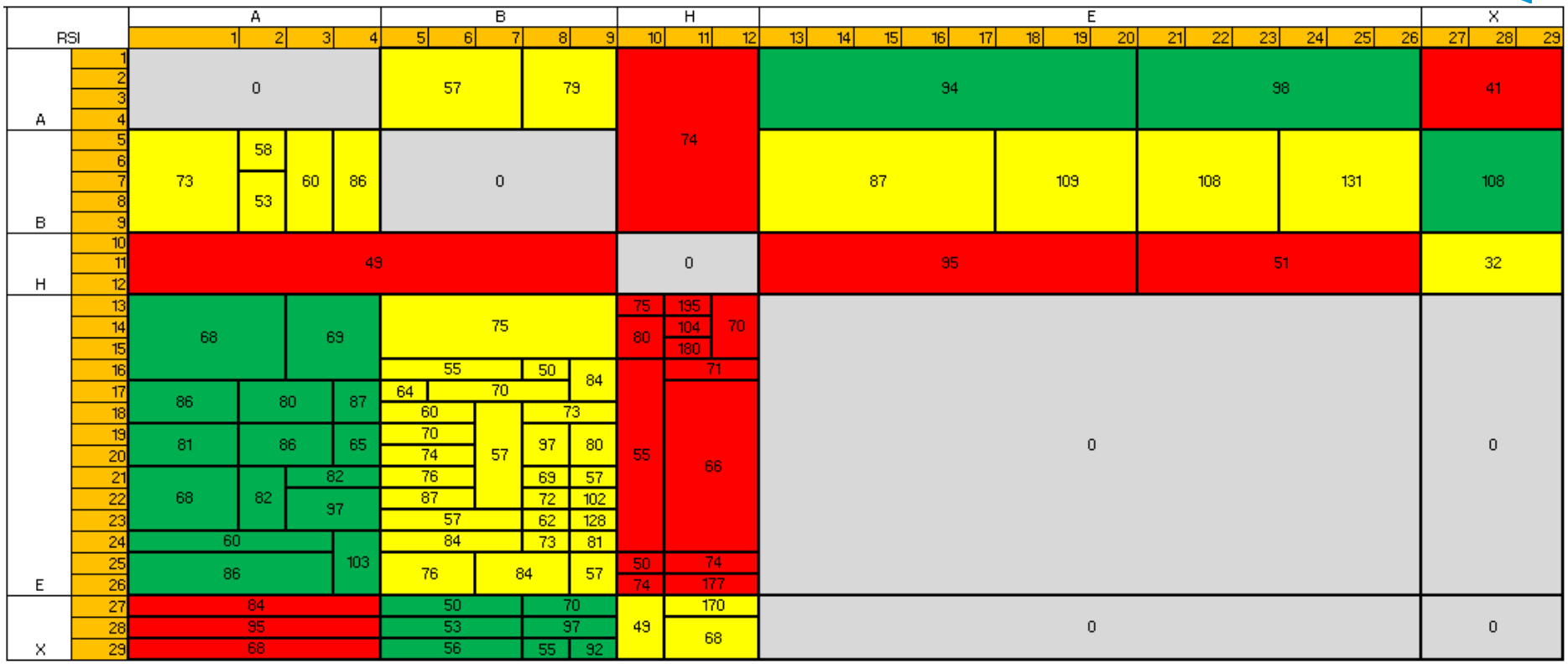
N.B. Zones outside the study area are defined as Area 'X'

Data processing

- Only AM processed
- Develop 29 sectors
- Sector cells aggregated into 184 ‘fusion blocks’ for statistical significance:
  - for observed RSI [sample Nij =50-100]
  - $var(N_{ij}) = N_{ij}.(1–N_{ij}/N)$
  - for unobserved [expanded Tij =300-400]

29 sectors, 184 fusion blocks

(Sample sizes for *observed* areas)



GPS or mobile for OD ?

GPS (Trafficmaster TM)

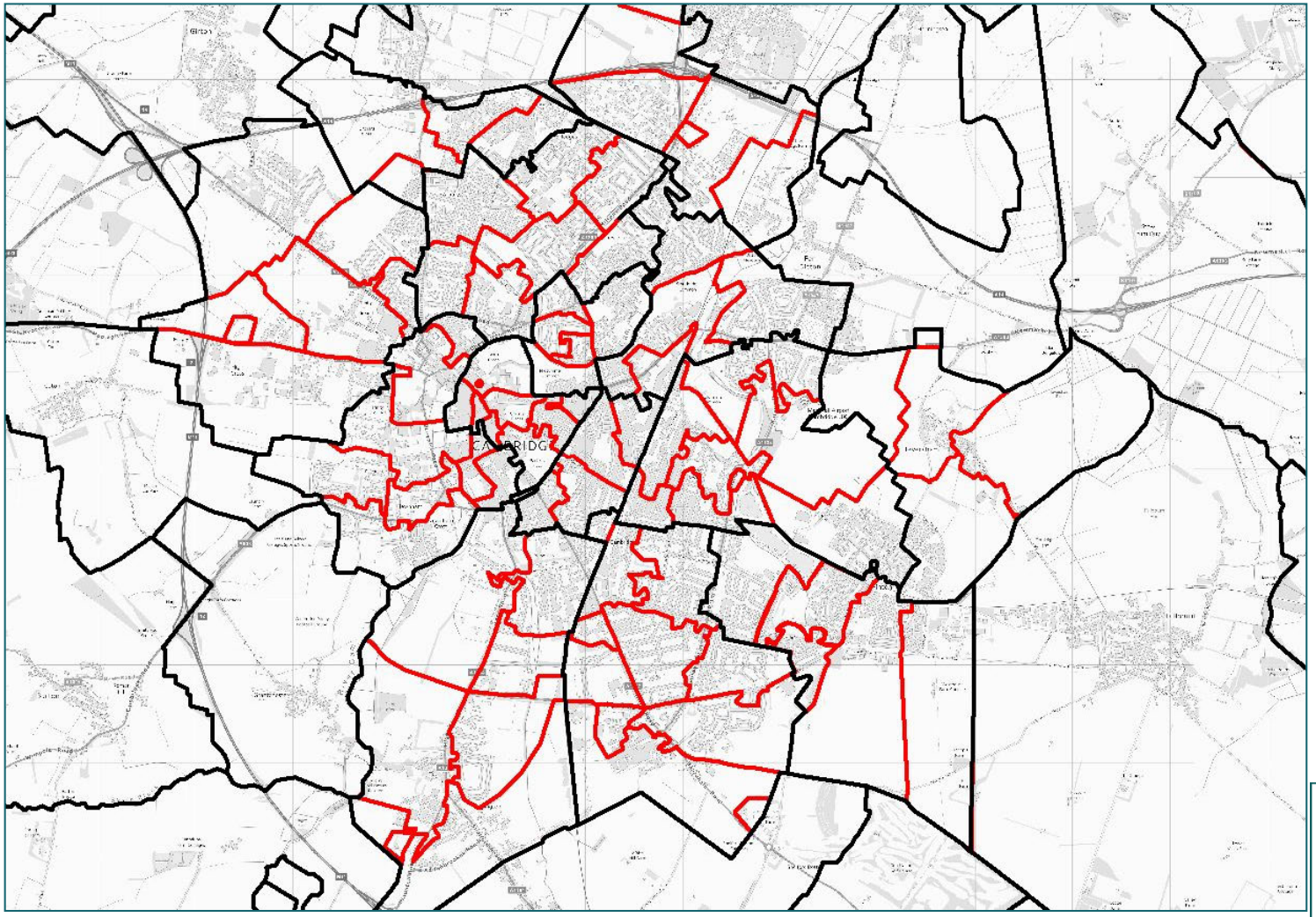
- ✓ **Spatial accuracy**
- ✓ Can track routes
- ✓ Availability / cost (TM is free for simple ODs)
- ✓ **Car, LGV, HGV**
- ❑ Only fixed units useful
- ❑ Bias dangers
- ❑ Low sample [c 0.5%] though easy to extend over time

Mobile (O2 thru INRIX)

- ✓ **Large sample** – up to c 28%
- ✓ Can track routes – at significant cost
- ✓ Little bias
- ❑ Spatial detail to *approx* XXX X postcode ***over a defined region***
- ❑ **Mode unknown** !

All big data lacks segmentation

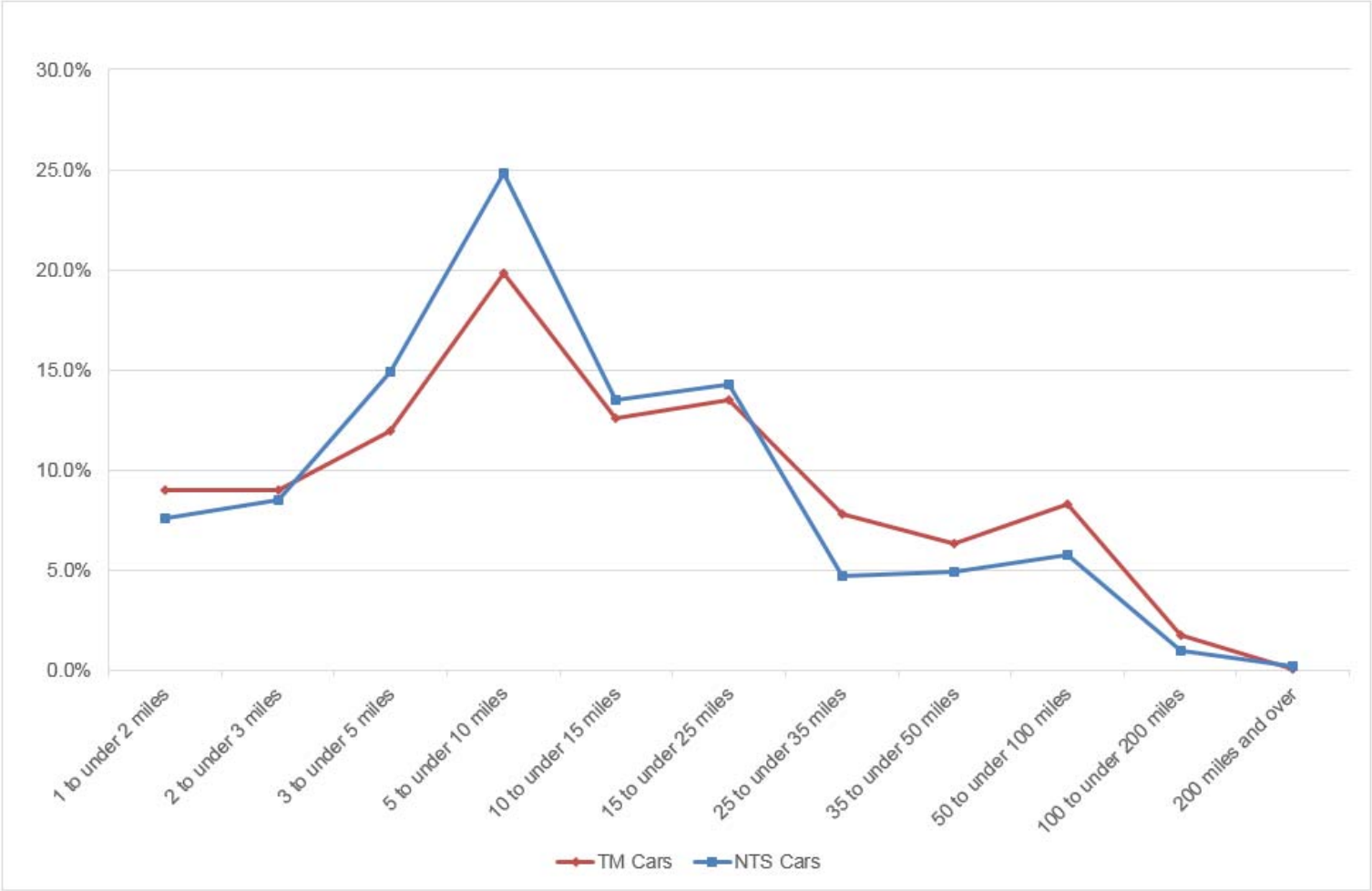
Mobile phone granularity



TM Data processing

- OD symmetry correlation R^2 0.99
- For car ownership R^2 0.94
- Correct for TLD

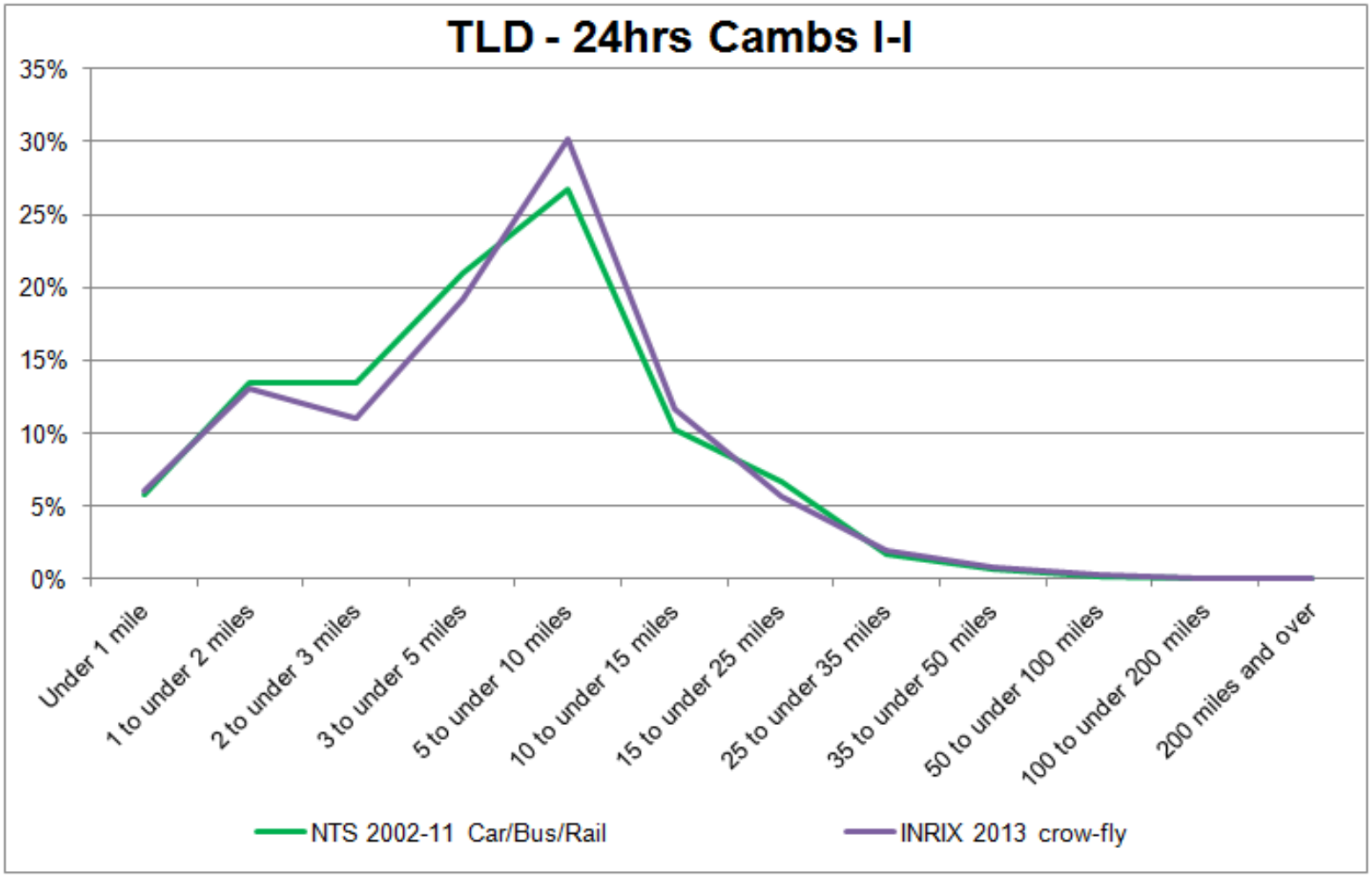
TLD for TM


30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
1 to under 2 miles
2 to under 3 miles
3 to under 5 miles
5 to under 10 miles
10 to under 15 miles
15 to under 25 miles
25 to under 35 miles
35 to under 50 miles
50 to under 100 miles
100 to under 200 miles
200 miles and over
TM Cars
NTS Cars

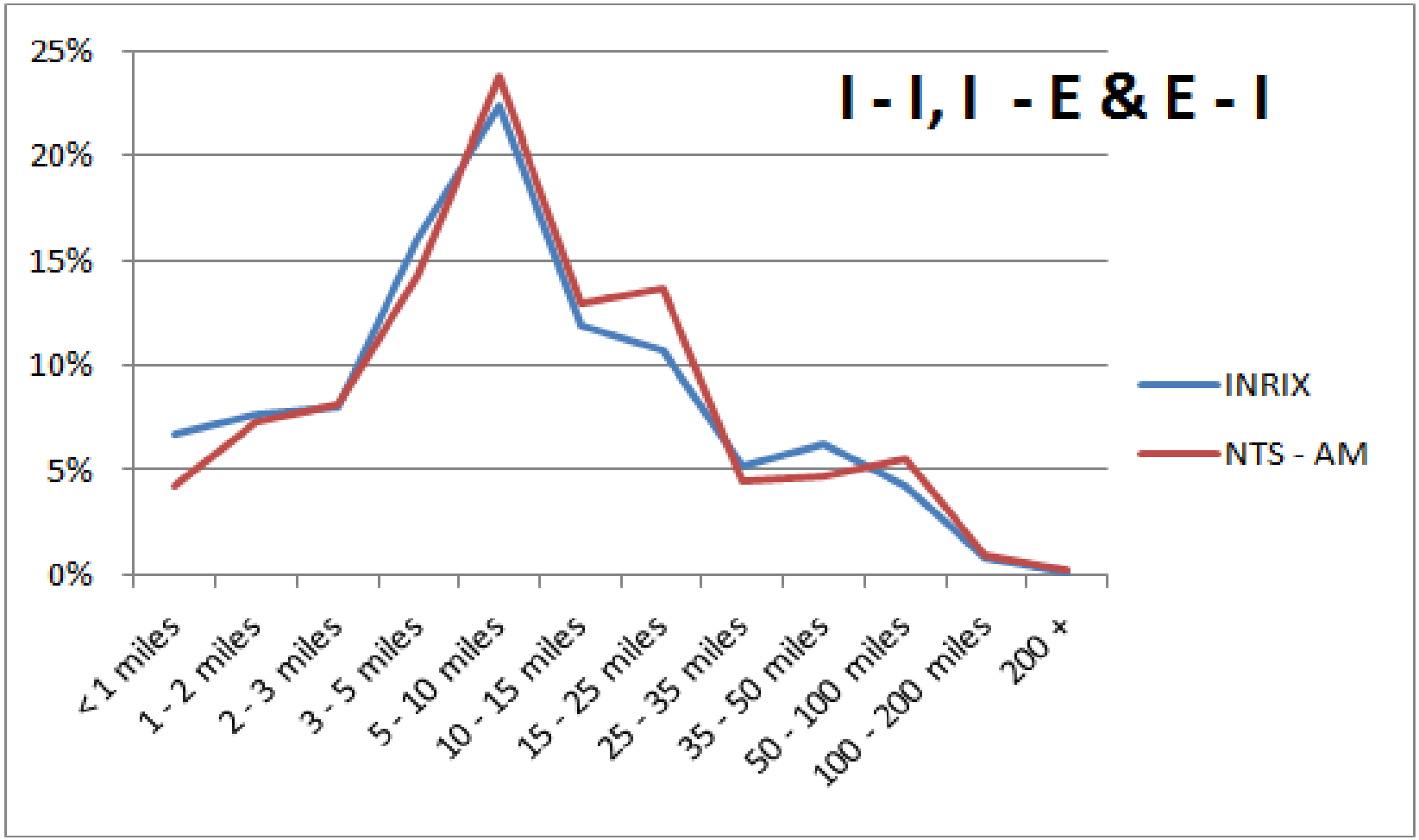
INRIX Data processing

- Pedestrian and cycle on speed criteria
- Symmetry test R 0.99
- Correction in several stages:
  - Check all-modes TLD
  - Correct for public transport at 29 sector level
  - Remove HGV using CSRM
  - Check & correct light veh. TLD

TLD – INRIX for All-Modes



TLD - INRIX Light Veh.s



Data expansion

- Data fused on proportions
- Expansion options:
    1. Global factor
    2. As 1, adjusted by observed to assigned
    3. Factor at 5 super sector
    4. As 3, constrain external to prior
    5. As 3, adjusted by observed to assigned

(Un)reliability of Inputs

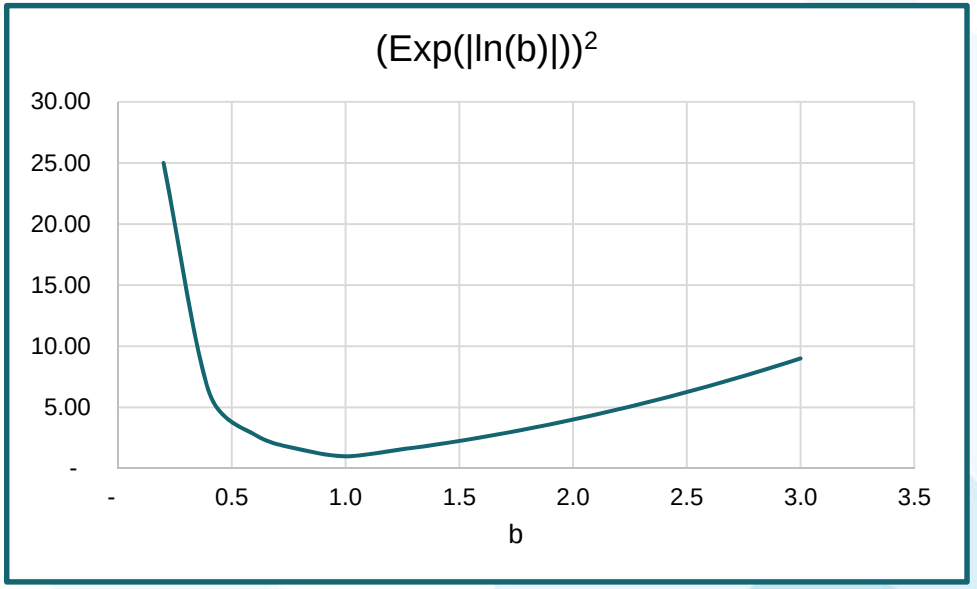
- Variance Prior Matrix RSI

$$V = p\,(1-p)/N$$

- Variance Prior Matrix un-observed

$$\text{Ln } V = a + b \text{ Ln } T$$

- Variance TM data
- INRIX additional factor



Covariances

$$\Omega_{IJ,RS} = \rho_{IJ,RS} \cdot \sqrt{\Omega_{IJ,IJ} \cdot \Omega_{RS,RS}}$$

$$\rho_{IJ,RS} = \frac{-p_{IJ} \cdot p_{RS}}{\sqrt{p_{IJ} \cdot p_{RS} \left(1 - p_{IJ}\right) \cdot \left(1 - p_{RS}\right)}}$$

Variables tested

- New data added
- Expansion method
- Variance of unobserved cells
- Inter-fusion block constraints
- Intra-fusion block constraints

Matrix Fusion Calculation

$$\boldsymbol{t'} = \left(\Omega_{\boldsymbol{t_1}}^{-\mathbf{1}} + \Omega_{\boldsymbol{t_2}}^{-\mathbf{1}}\right)^{-\mathbf{1}} \left(\Omega_{\boldsymbol{t_1}}^{-\mathbf{1}} \boldsymbol{t_1} + \Omega_{\boldsymbol{t_2}}^{-\mathbf{1}} \boldsymbol{t_2}\right)$$

t_1 and t_2 – input demand datasets
Ω_1 and Ω_2 – associated covariance matrices

Computationally very difficult
Can be simplified considerably!

Define success

- Statistical uncertainty of fused proportions
- Assignment validation of updated prior

- Scale of change by ME
- Assignment validation of updated post ME
- Ease of use by practitioner

Test programme

ID		Prior Matrix Variance Assumptions for Fusion Blocks		Processing Assumptions		
Run ID	Data to be fused with CSRM Prior	Variance of EERM-sourced prior data	Variance of other unobserved prior data	Source of constraints for inter-fusion block unobserved movements	Source of Intra-Fusion Block Distribution for use in output matrix	Expansion method for 184 Fusion Block proportions from MATLAB
18	TM	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	Prior	3
19	TM	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	Prior	4
22	INRIX	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	Prior	3
23	INRIX	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	Prior	4
24	TM & INRIX	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	Prior	3
25	TM & INRIX	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	Prior	4
26	TM	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	TM	1
27	TM	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	TM	2
28	TM	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	TM	5
29	INRIX	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	TM	1
30	INRIX	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	TM	5
31	TM & INRIX	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	TM	1
33	INRIX	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	INRIX	1
34	INRIX	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	INRIX	5
35	TM	4*'RSI-Eq.'	4*'RSI-Eq.'	Matrix Fusion	TM	5
36	TM	2*'RSI-Eq.'	½*'RSI-Eq.'	Matrix Fusion	TM	5
39	TM	N/A	N/A	TM	TM	5
40	TM	N/A	N/A	Prior	TM	5
41	TM	N/A	N/A	(TM+Prior)/2	TM	5
42	INRIX	N/A	N/A	Prior	INRIX	5
43	INRIX	N/A	N/A	(INRIX+Prior)/2	INRIX	5
44	TM	2*'RSI-Eq.'	2*'RSI-Eq.'	Matrix Fusion	Prior	5

Test results

- New data sources certainly improve prior
- Simple inverse variance weighting preferable
- Expand at 5 super sector level, adjusted by observed to assigned
- Best results from adding INRIX data using TM distribution
- Fusion of un-observed cells not justified, take mean of prior and new data


Q&A