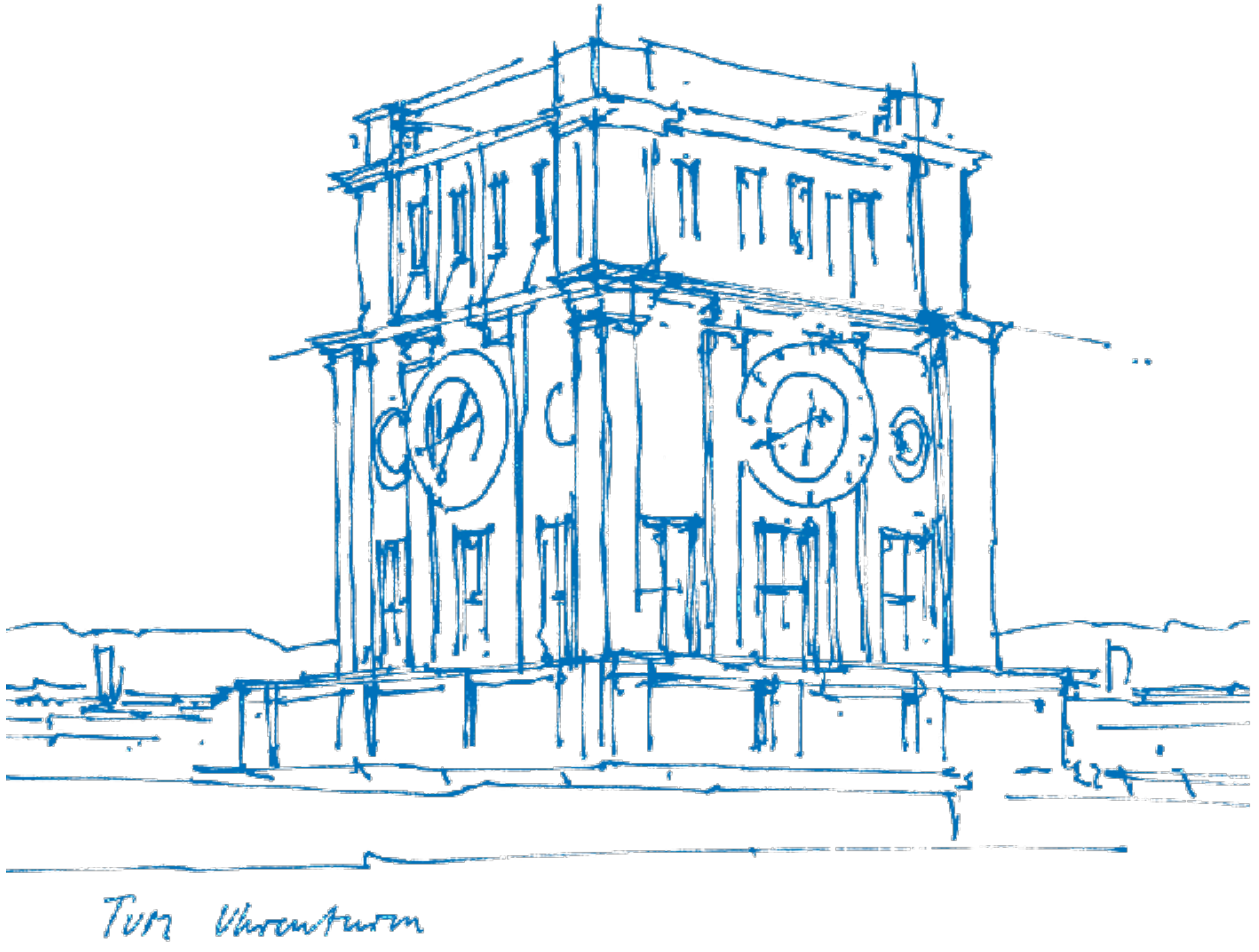
Big data for transportation systems analysis
Challenges and opportunities

Univ.-Prof. Dr. Constantinos Antoniou
Technische Universität München
Fakultät für Bau, Geo und Umwelt
Lehrstuhl für Vernetzte Verkehrssysteme
Graz, 17. Mai 2018



Outline

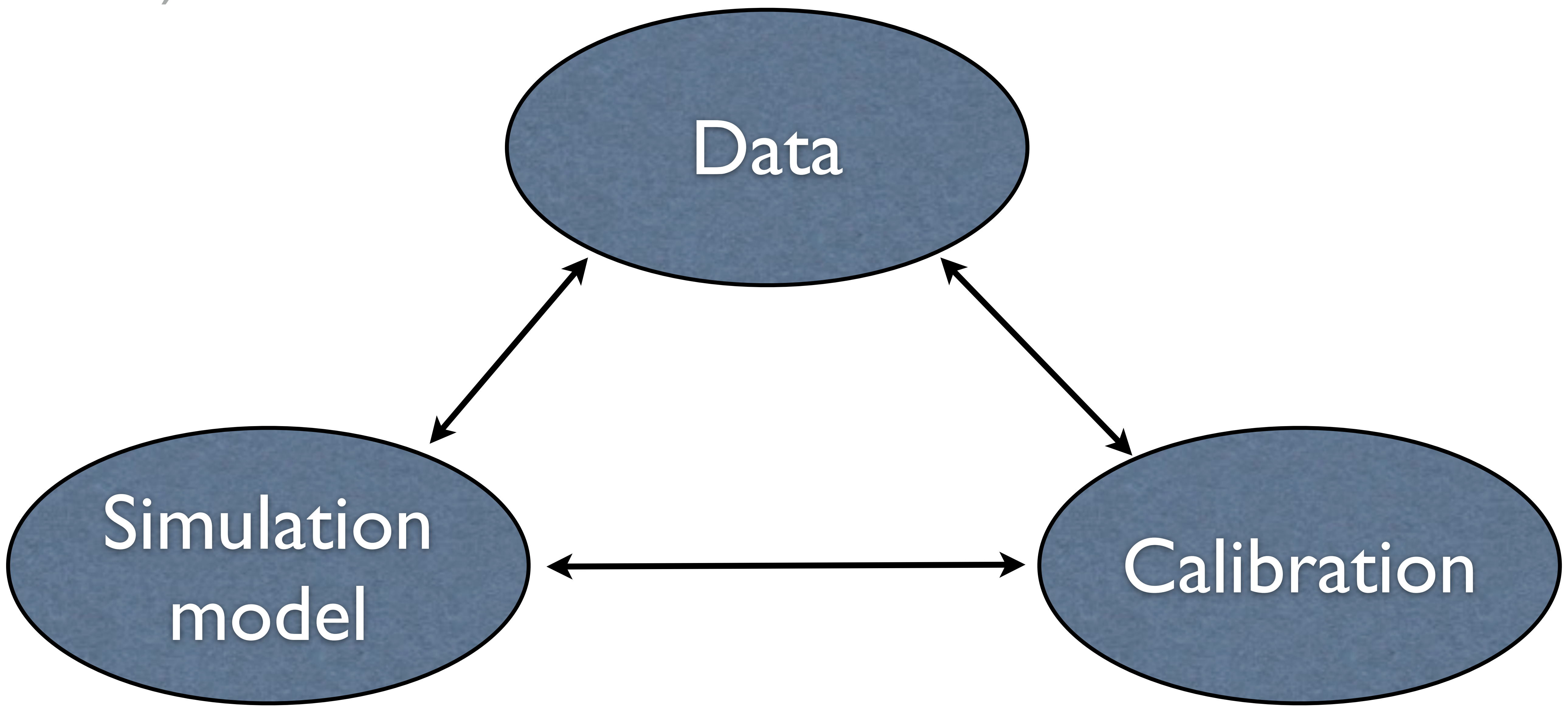
Transportation systems

Big data - what they are (and what they are not)

Opportunities

Challenges

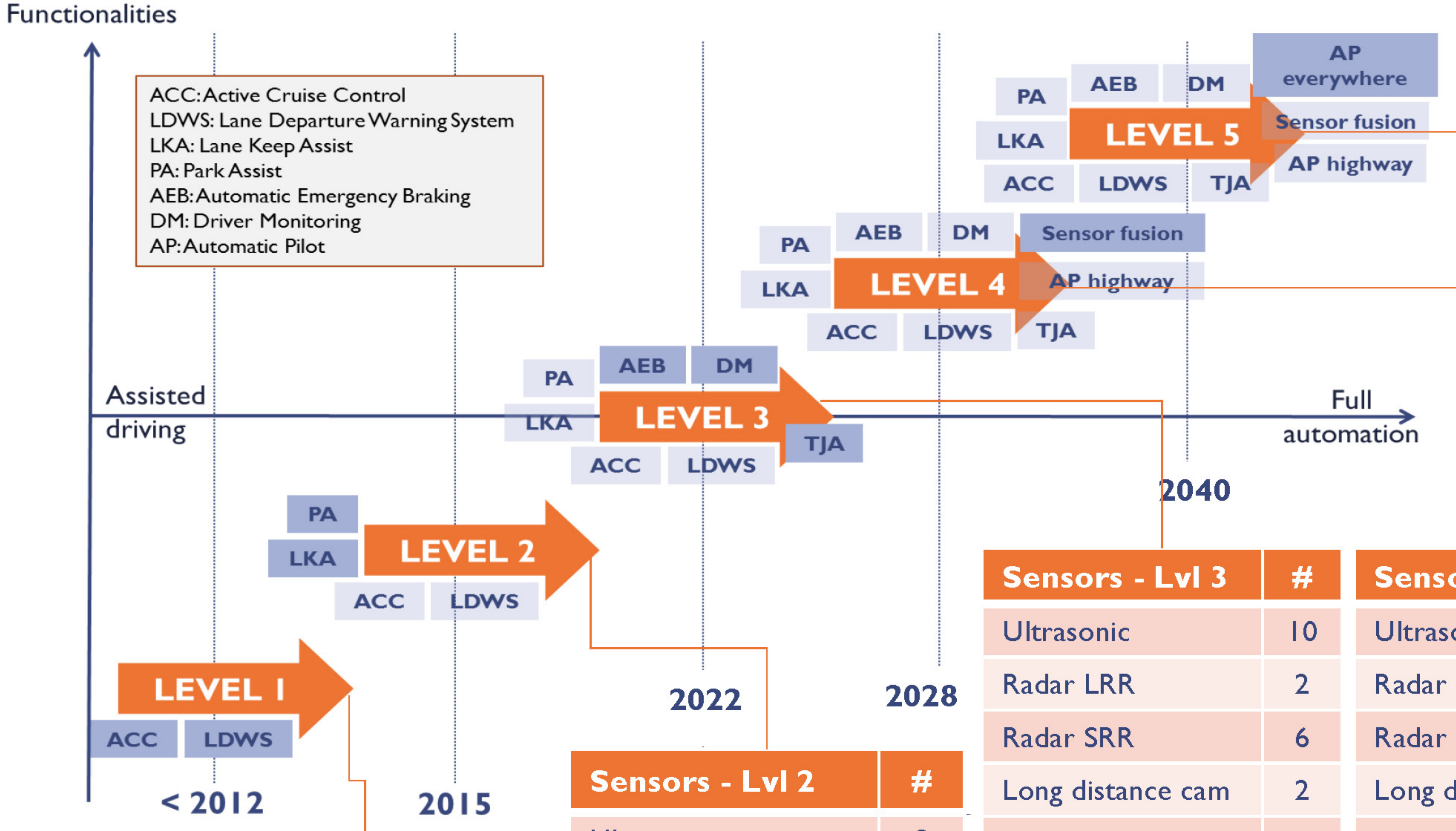
Outlook



Lehrstuhl für Vernetzte Verkehrssysteme
Fakultät für Bau, Geo und Umwelt
Technische Universität München

SENSOR TECHNOLOGY ROADMAP AND AUTONOMOUS FUNCTIONS ASSOCIATED

(Source: Sensors & Data Management for Autonomous Vehicles report, Oct. 2015, Yole Développement)



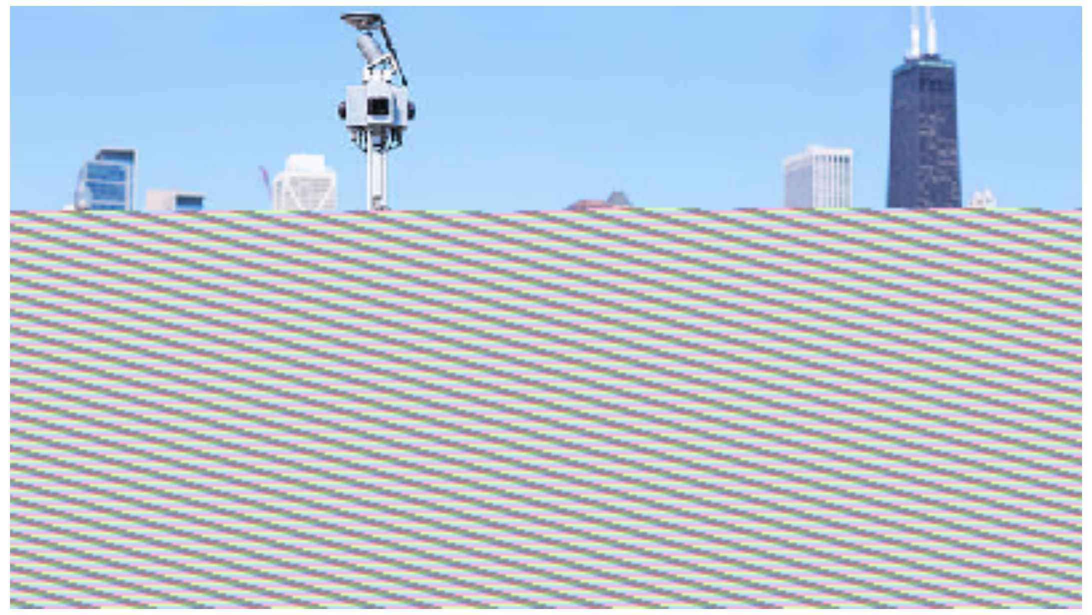
Sensors - Lvl 1	#
Ultrasonic	4
Radar LRR	1
Camera for surround	1
TOTAL	**6**

Sensors - Lvl 2	#
Ultrasonic	8
Radar LRR	1
Radar SRR	4
Camera for surround	4
TOTAL	**17**

Sensors - Lvl 3	#
Ultrasonic	10
Radar LRR	2
Radar SRR	6
Long distance cam	2
Camera surround	5
Stereo camera	1
µbolo	1
LIDAR	1
Dead reckoning	1
TOTAL	**29**

Sensors – Lvl 4	#
Ultrasonic	10
Radar LRR	2
Radar SRR	6
Long distance cam	2
Camera surround	5
Stereo camera	1
µbolo	1
LIDAR	1
Dead reckoning	1
TOTAL	**29**

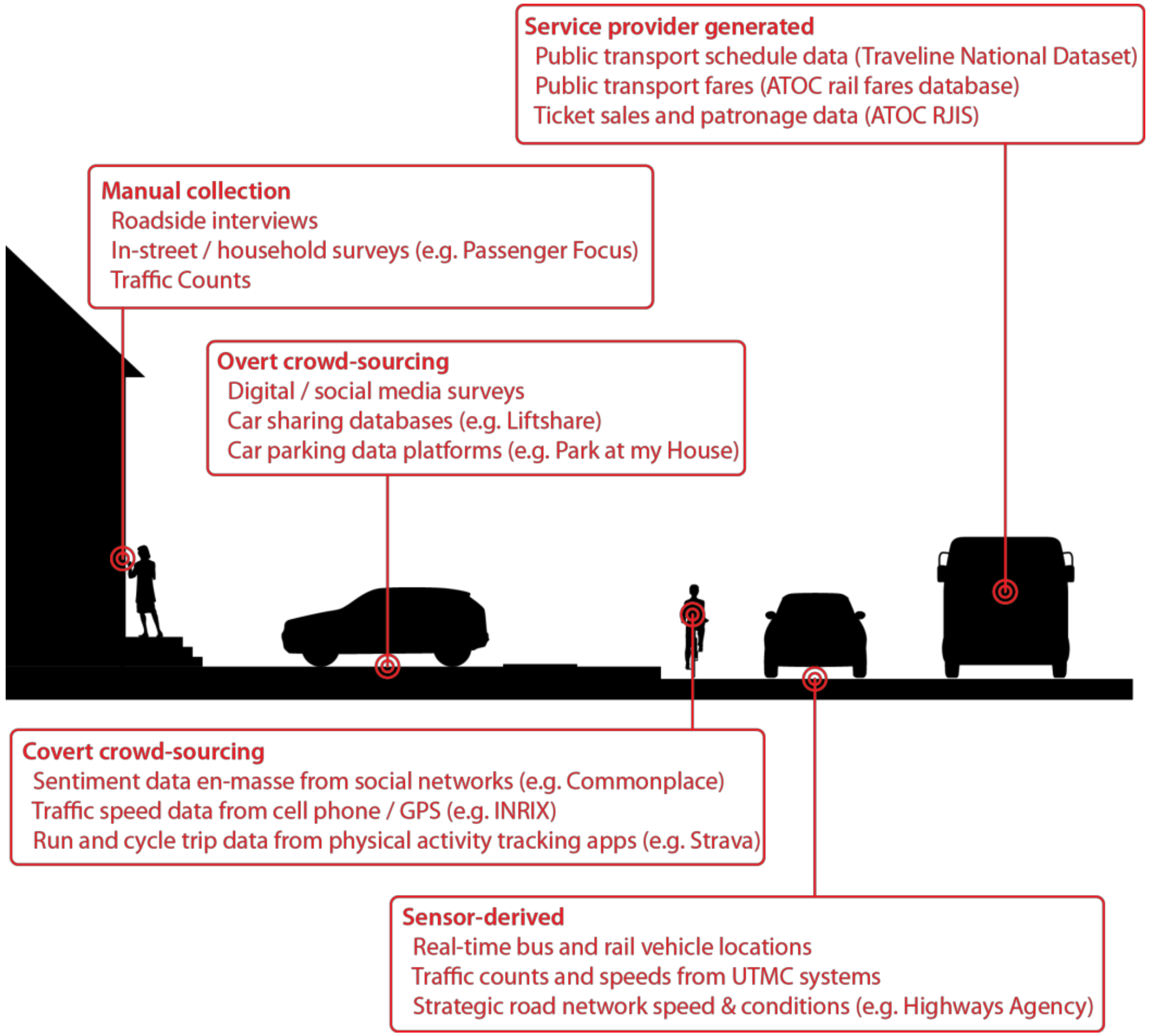
Sensors – Lvl 5	#
Ultrasonic	10
Radar LRR	2
Radar SRR	6
Long distance cam	4
Camera surround	5
Stereo camera	2
µbolo	1
LIDAR	1
Dead reckoning	1
TOTAL	**32**



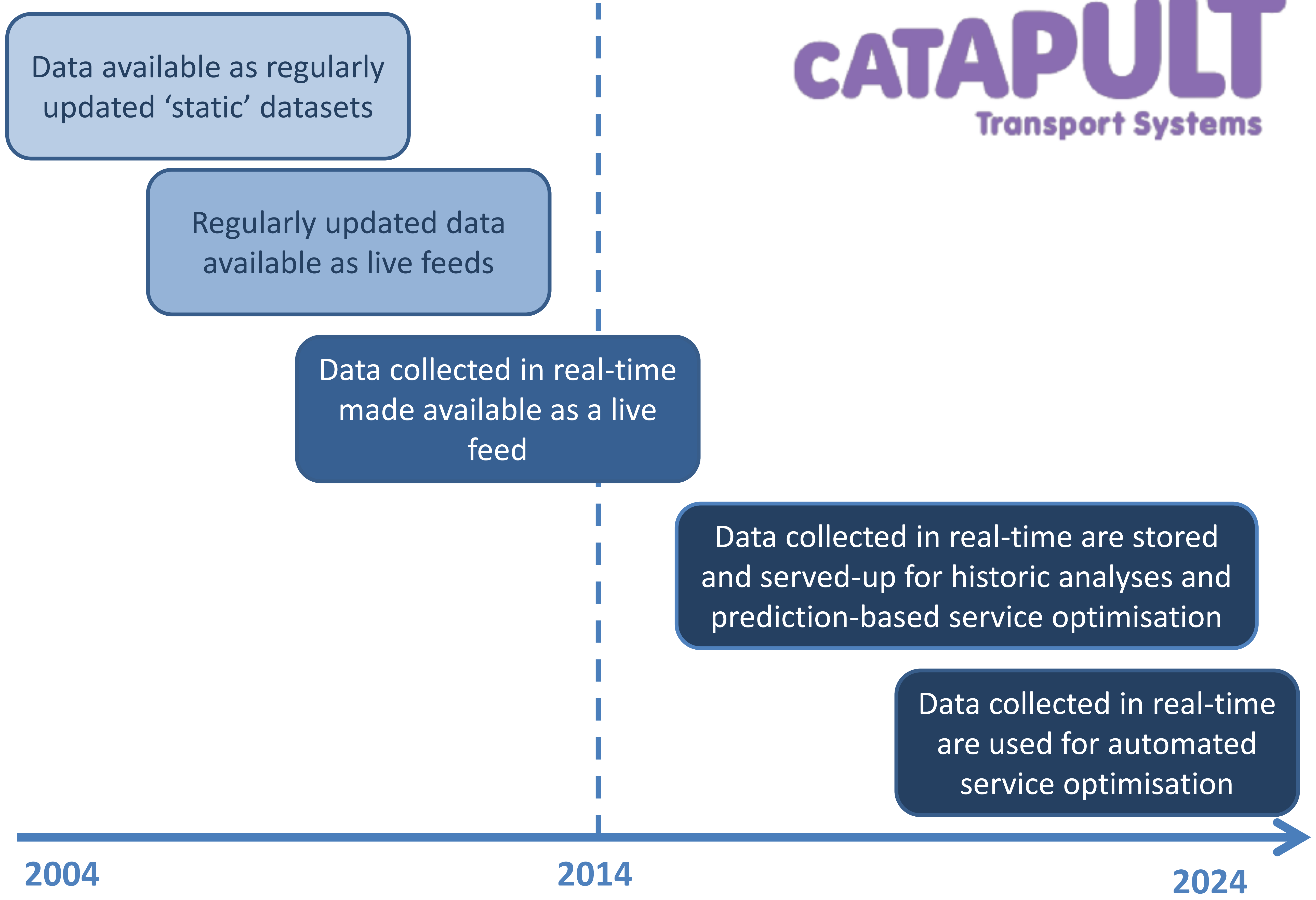
©2015 | www.yole.fr

Lehrstuhl für Vernetzte Verkehrssysteme
Fakultät für Bau, Geo und Umwelt
Technische Universität München

Transportation systems data



CATAPULT Transport Systems



But what is Big Data? (The 3Vs)

Volume:
Increasingly massive datasets hard to manage

Large Hadron Collider experiment, 150 million sensors delivering data 40 million times per second.

Variety:
Data complexity is growing

More types of data captured than ever before, quantification of self etc.

But what is Big Data? (The 3Vs - cont'd)

Velocity:

Some data is arriving so rapidly it must be either processed instantly or lost

Whole subfield of 'streaming data'

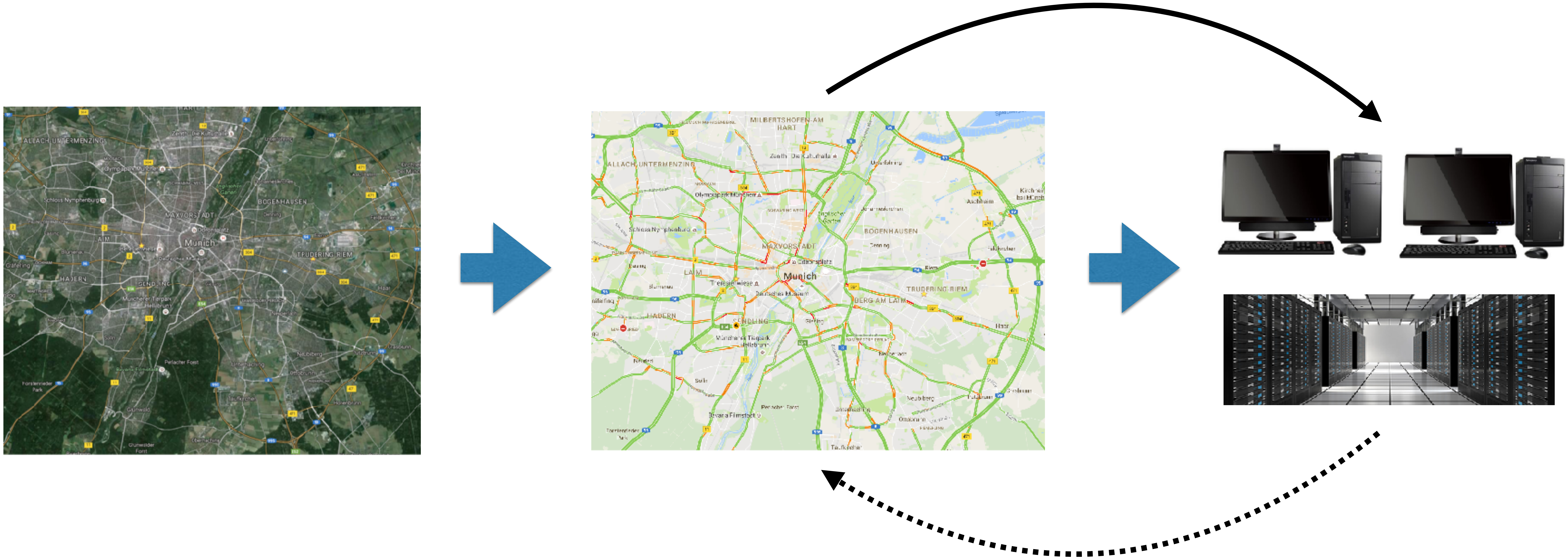
Veracity

Value: this is very important, this is really the output

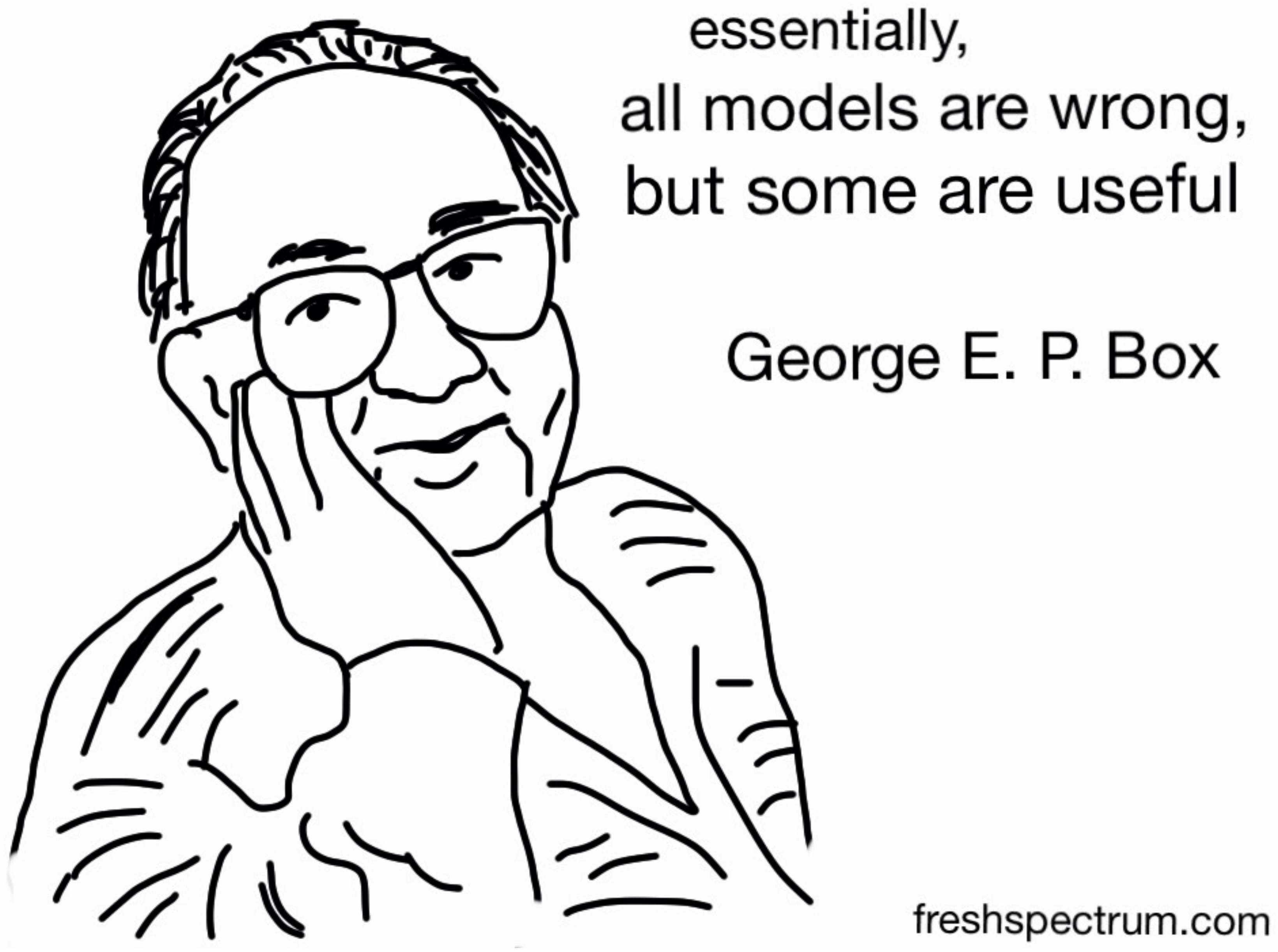

BIG DATA
— GOT IT...
NOW WHAT?

Lehrstuhl für Vernetzte Verkehrssysteme
Fakultät für Bau, Geo und Umwelt
Technische Universität München

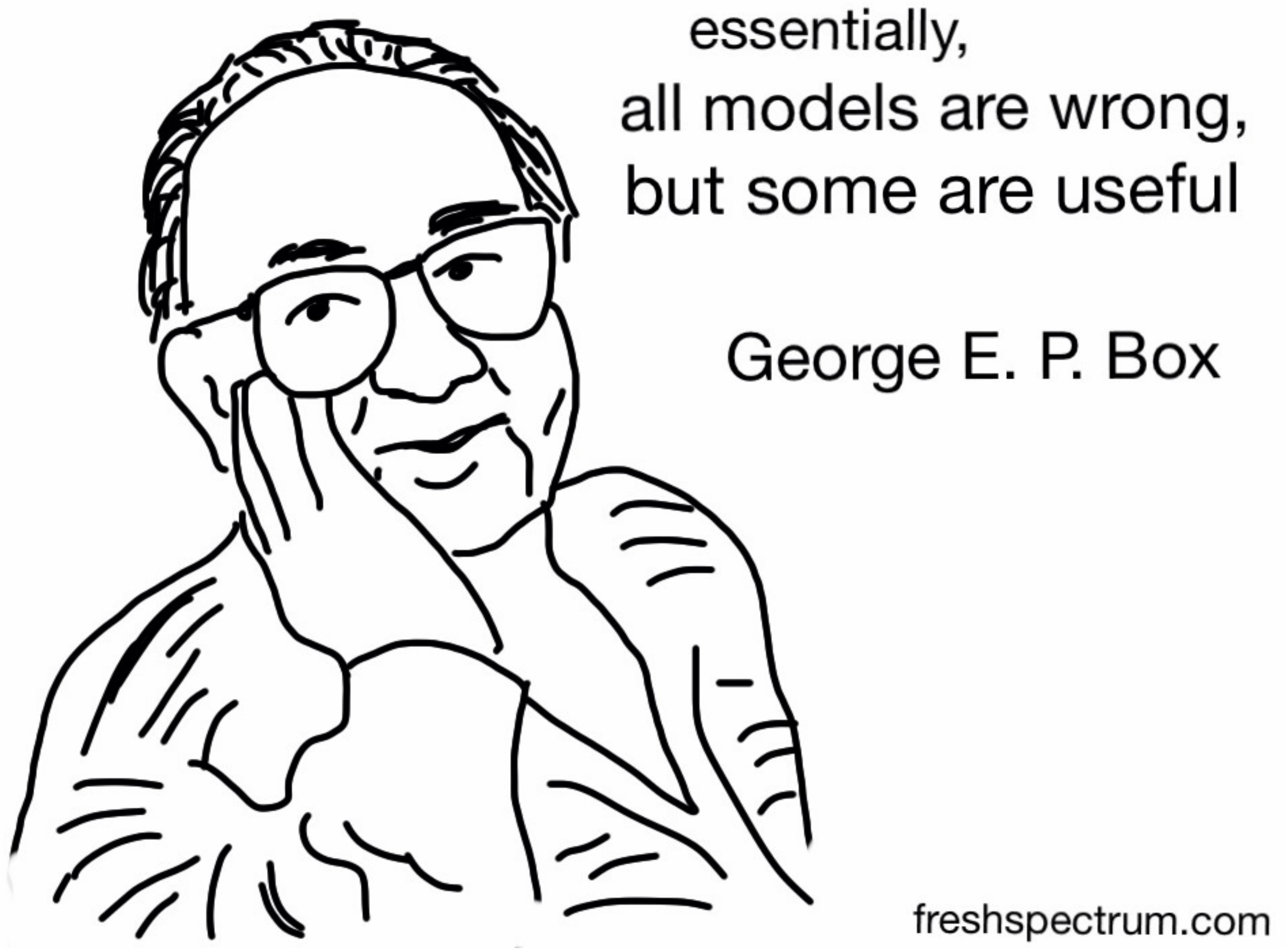
Historical modeling approach



Building a model

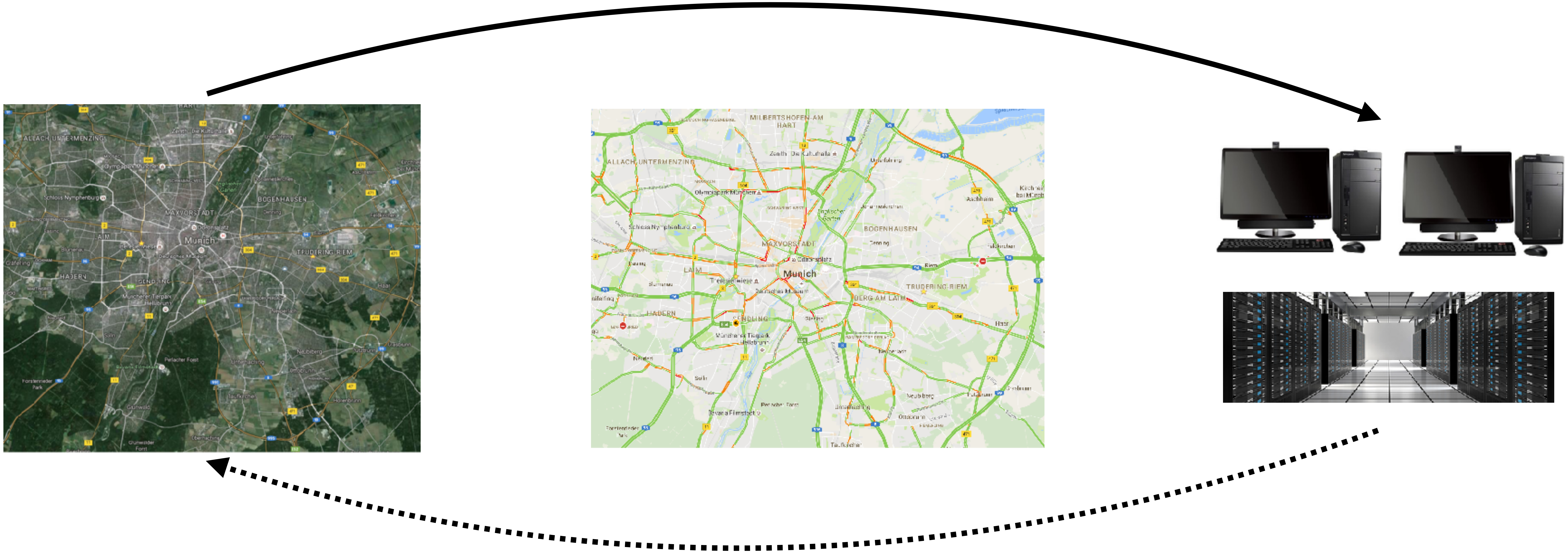


Building a model

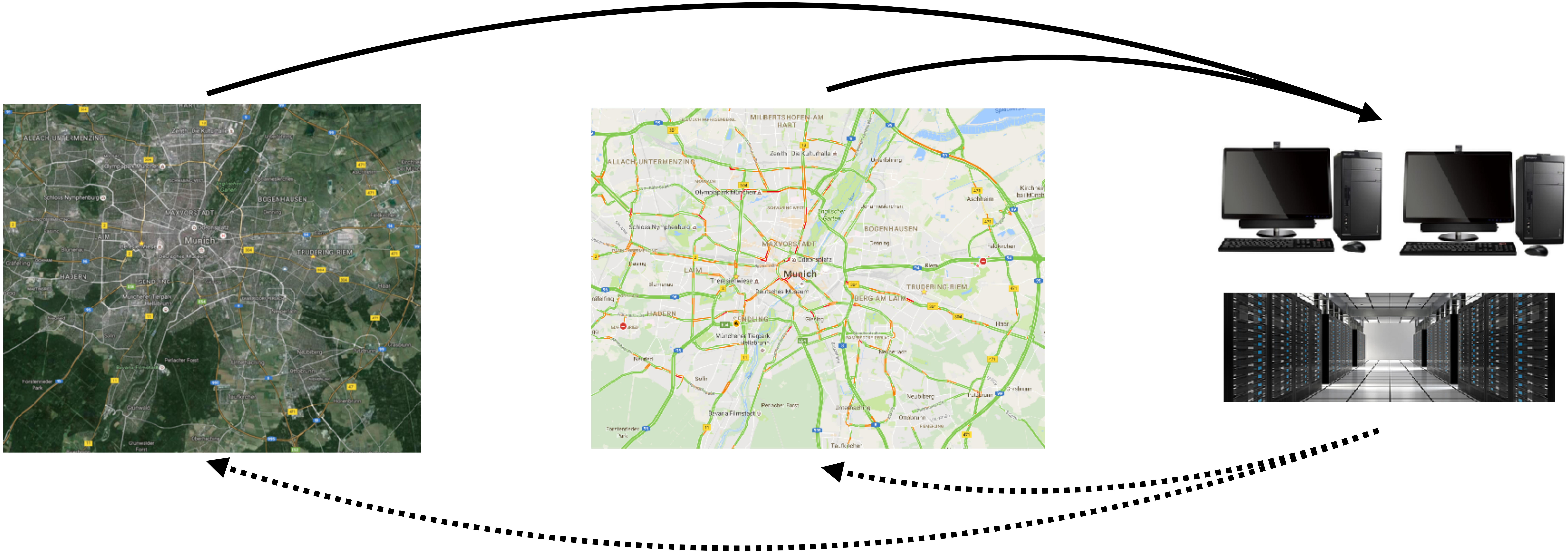


~~“All models are wrong,~~
~~and we can increasingly succeed without them”~~
~~C. Anderson (misquoting P. Norvig)~~
~~http://norvig.com/fact-check.html~~

First-generation big-data approach

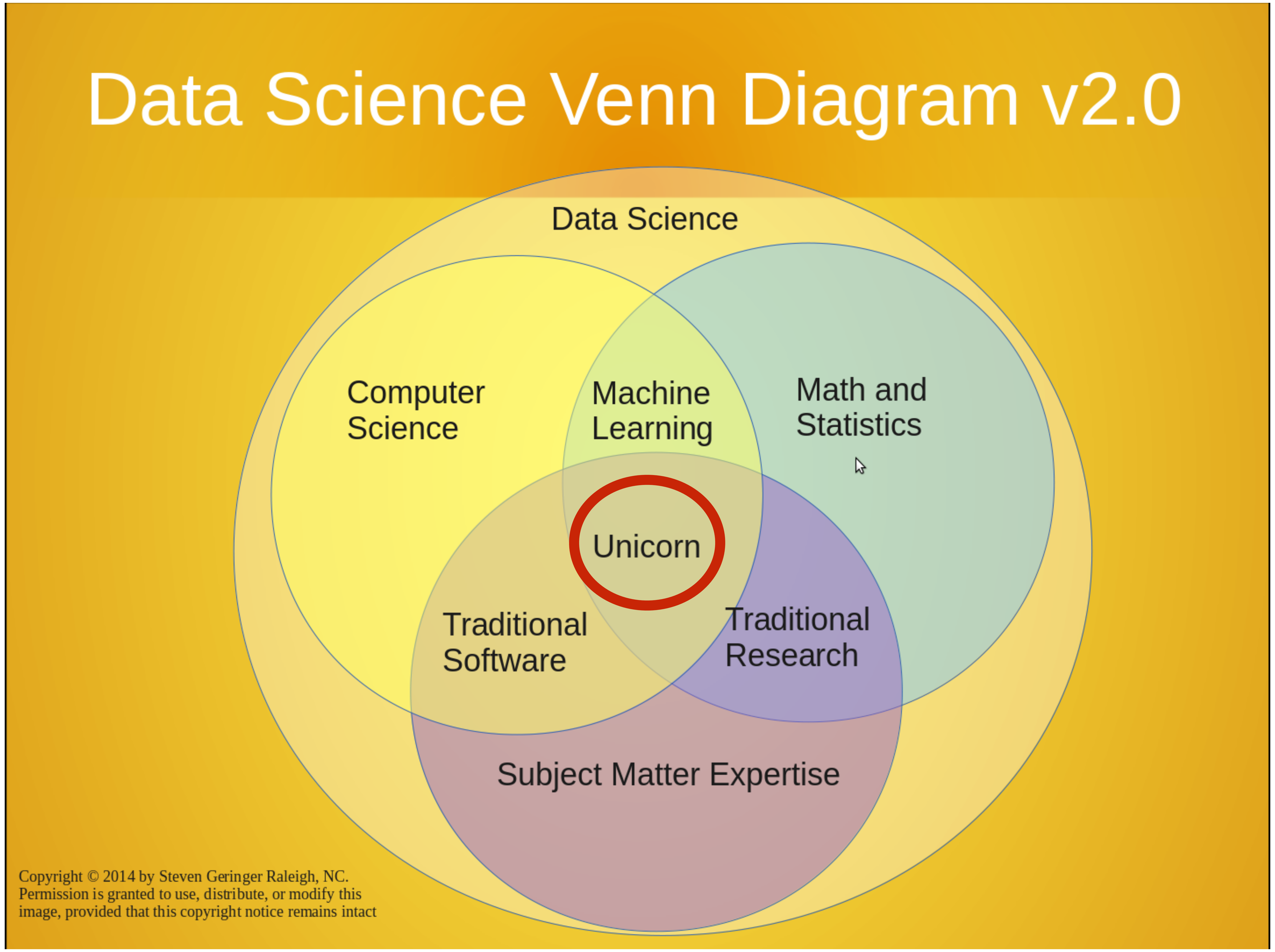


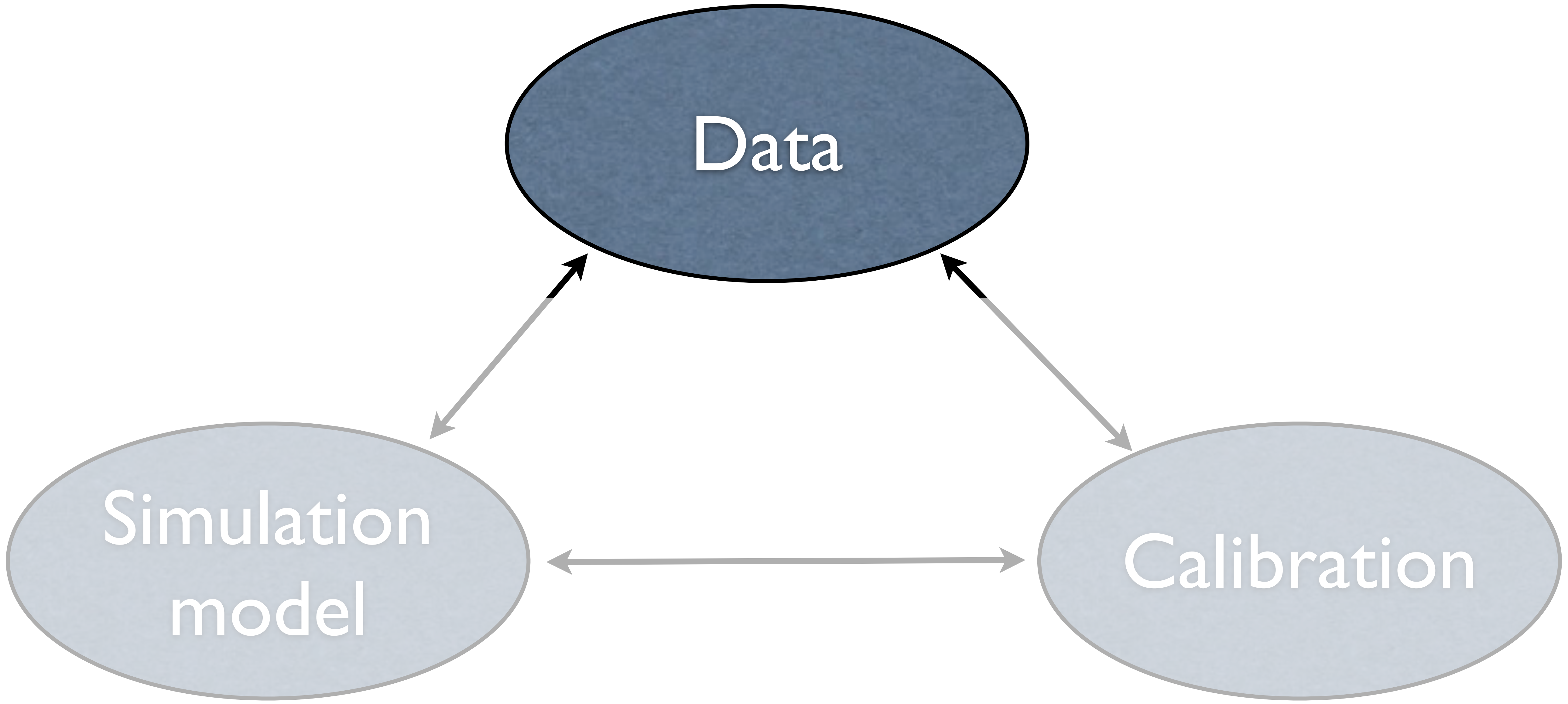
Second-generation big-data approach

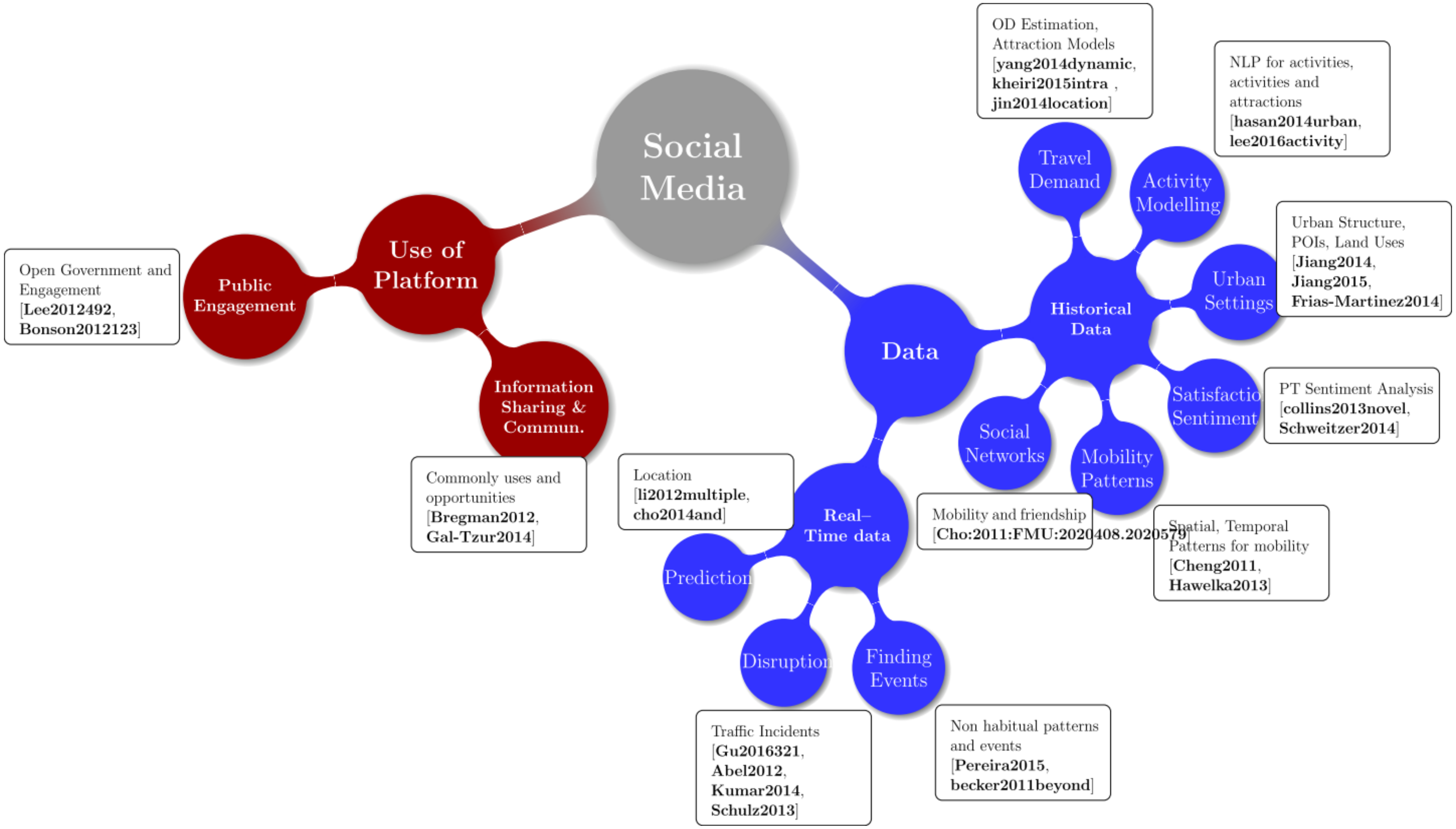


New skill-set is needed

(Look for the intersection in this Venn diagram!)




Data
Simulation model
Calibration



Chaniotakis, Antoniou, Pereira. "Mapping Social Media for Transportation Studies." *IEEE Intelligent Systems* 31.6 (2016): 64-70.

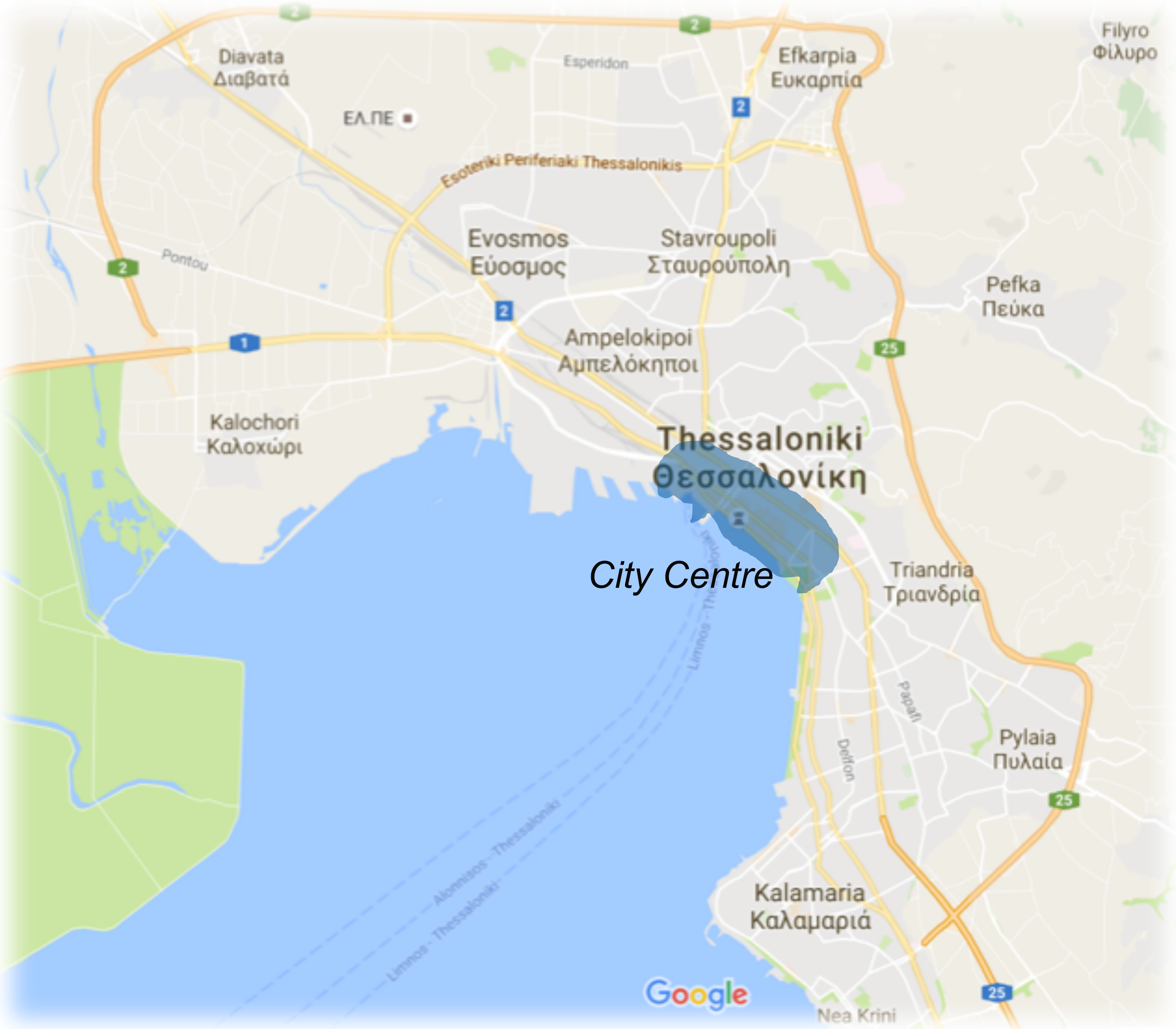
Case Study: Social Media vs. Surveys

Study Area
Thessaloniki, Greece
2nd largest city in Greece
Moderate Social Media use

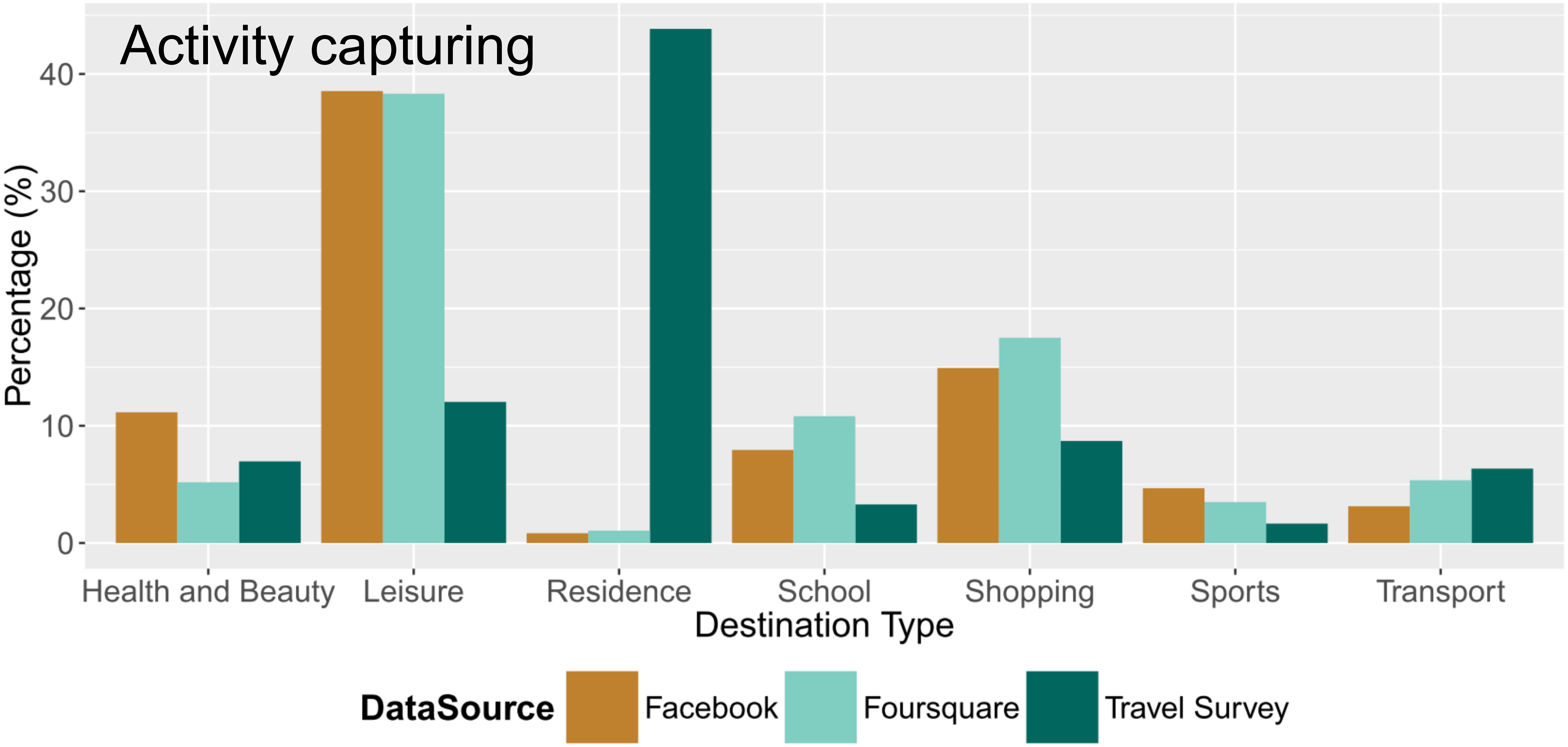
Datasets
Social Media
- Facebook
- Twitter
- Foursquare

Recent travel diary survey

FOURSQUARE

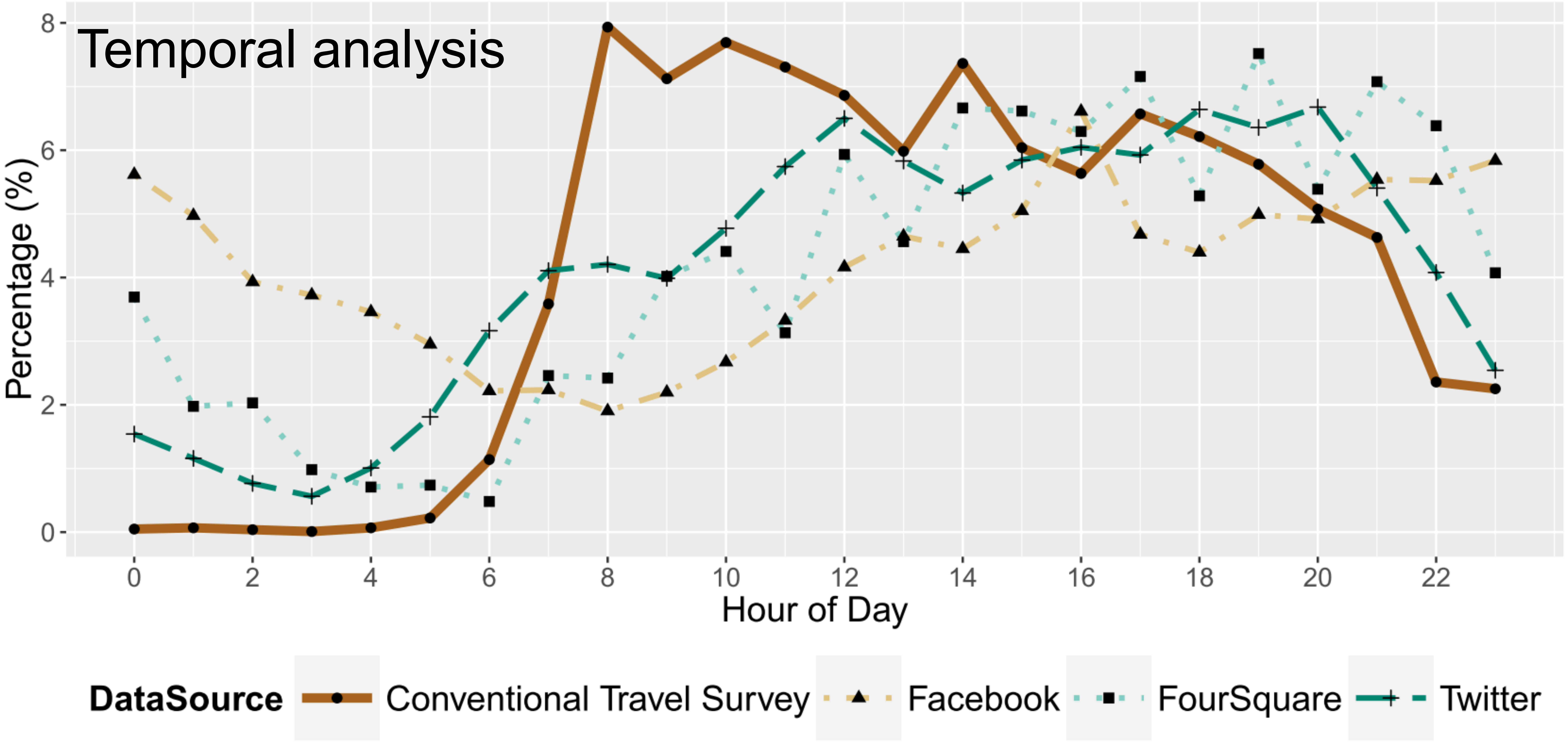


Chaniotakis, Antoniou, Salanova, Dimitriou, "Can Social Media data augment travel demand survey data?." *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016.



Chaniotakis, Antoniou, Salanova, Dimitriou, "Can Social Media data augment travel demand survey data?." *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016.

Temporal analysis



Chaniotakis, Antoniou, Salanova, Dimitriou, "Can Social Media data augment travel demand survey data?." *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016.

Lehrstuhl für Vernetzte Verkehrssysteme
Fakultät für Bau, Geo und Umwelt
Technische Universität München

Case study: Inferring activity types

Dataset

2 years' data collected from London (Twitter API)

482,883 unique users

Collected timeline (for a random sample of 90,000 users)

11,060,814 tweets in total

In London Twitter Dataset
482,883 Users
8,141,996 geotagged tweets
3,764,230 URLs
220,118 Foursquare URLs

{subset}
{*nTweet* > 22 & *mean dist* > 1}

Examined Dataset
50,344 Users
5,080,362 geotagged tweets
2,127,071 URLs
145,192 Foursquare URLs

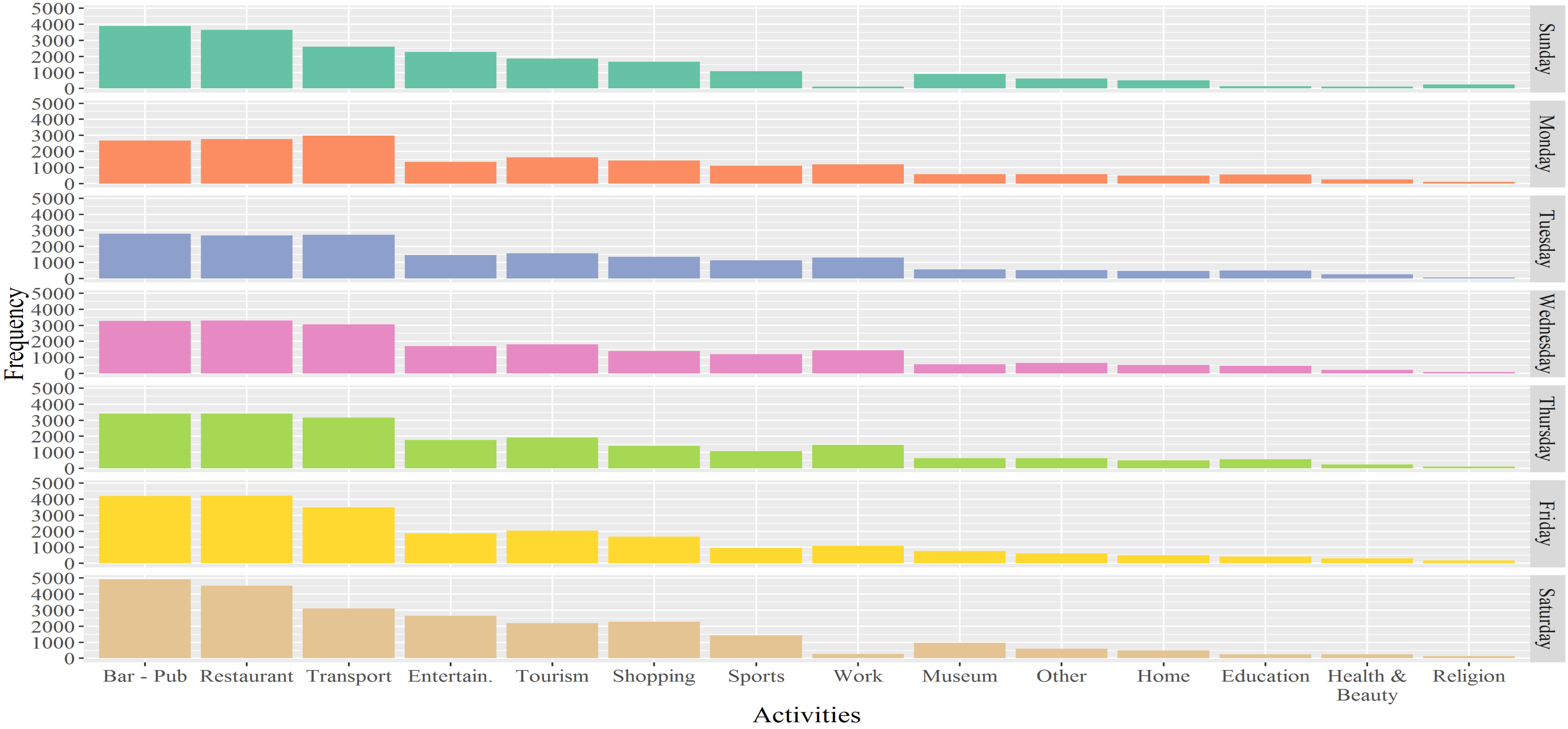
Chaniotakis, E., C. Antoniou, G. Aifadopoulou and L. Dimitriou (2017). Inferring activities from Social Media data. Transportation Research Record: Journal of the Transportation Research Board

Supervised training

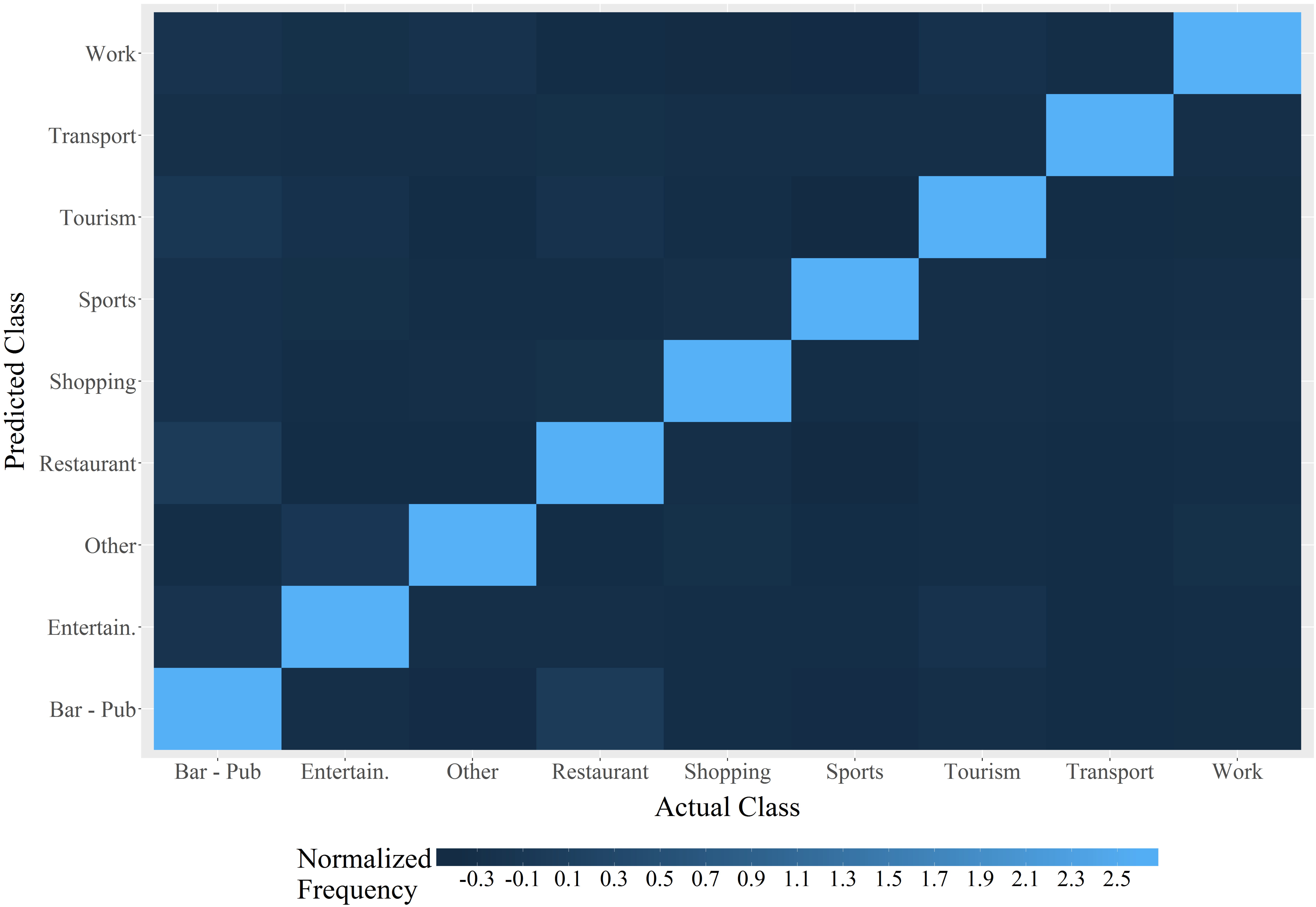
Manually aggregated in 14 categories

Tendency towards leisure activities

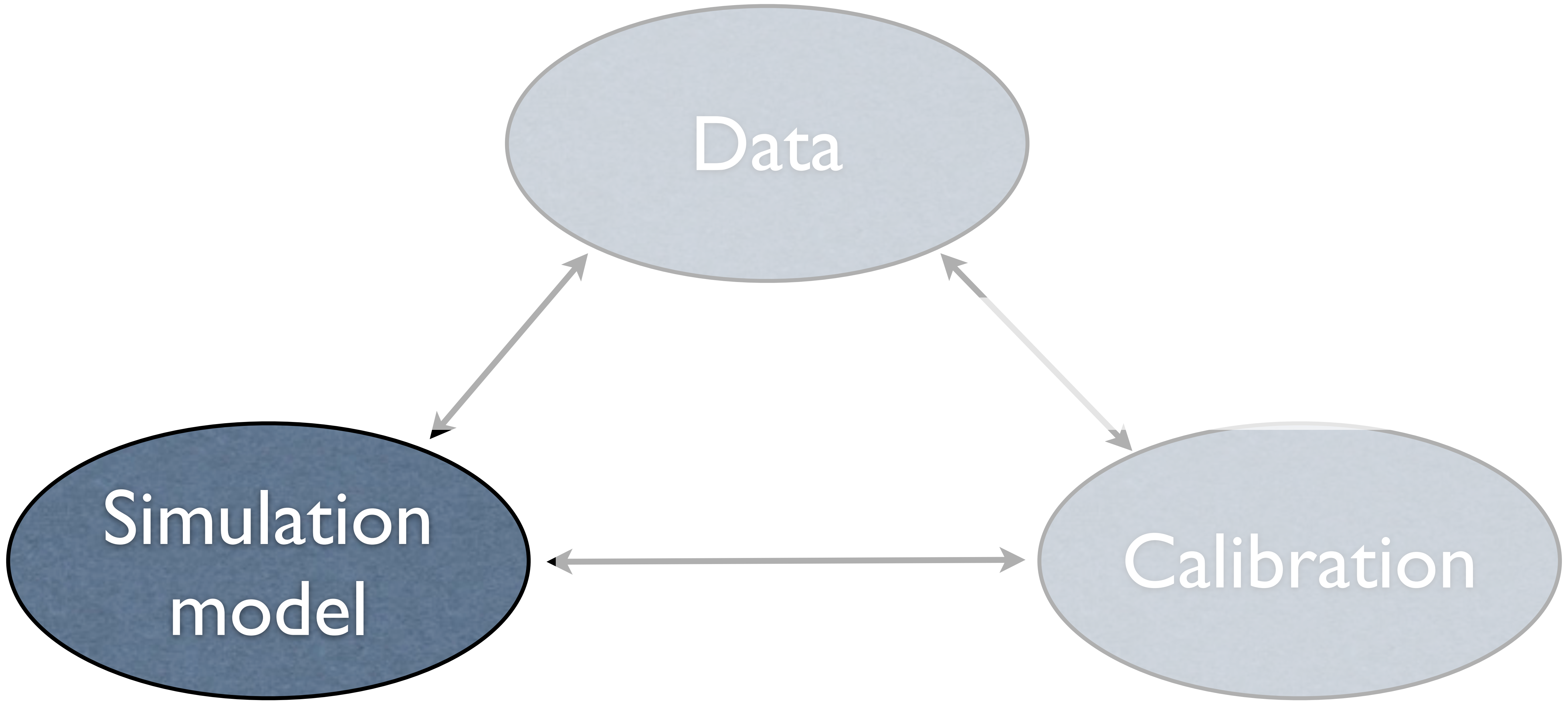
Education and work higher represented during week days



Chaniotakis, E., C. Antoniou, G. Aifadopoulou and L. Dimitriou (2017). Inferring activities from Social Media data. Transportation Research Record: Journal of the Transportation Research Board


Work
Transport
Tourism
Sports
Shopping
Restaurant
Other
Entertain.
Bar - Pub
Predicted Class
Bar - Pub
Entertain.
Other
Restaurant
Shopping
Sports
Tourism
Transport
Work
Actual Class
Normalized Frequency
-0.3 -0.1 0.1 0.3 0.5 0.7 0.9 1.1 1.3 1.5 1.7 1.9 2.1 2.3 2.5

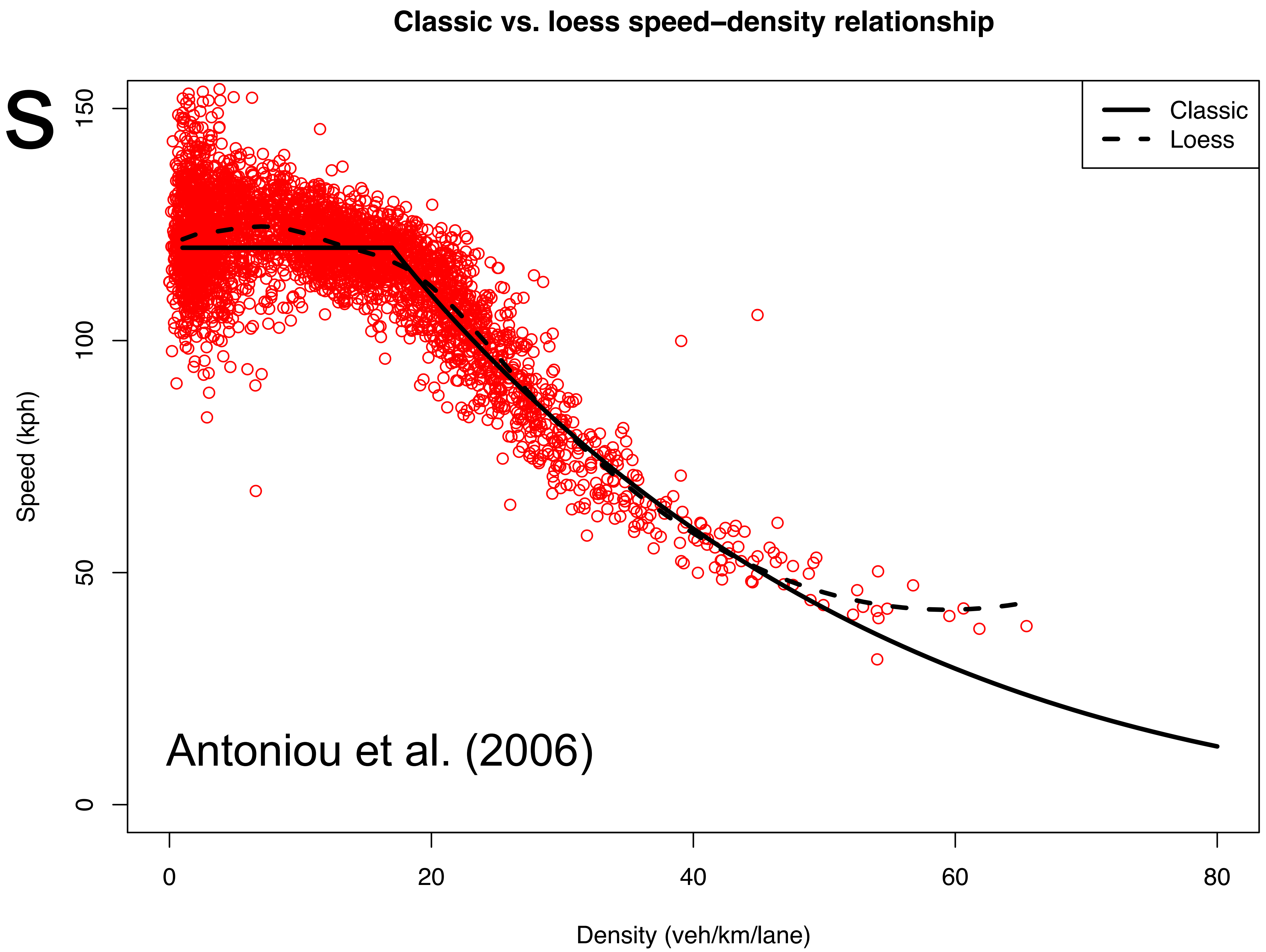
Max Entropy


Data
Simulation model
Calibration

Flexible traffic simulation models

Flexible functional form

Ability to incorporate additional data

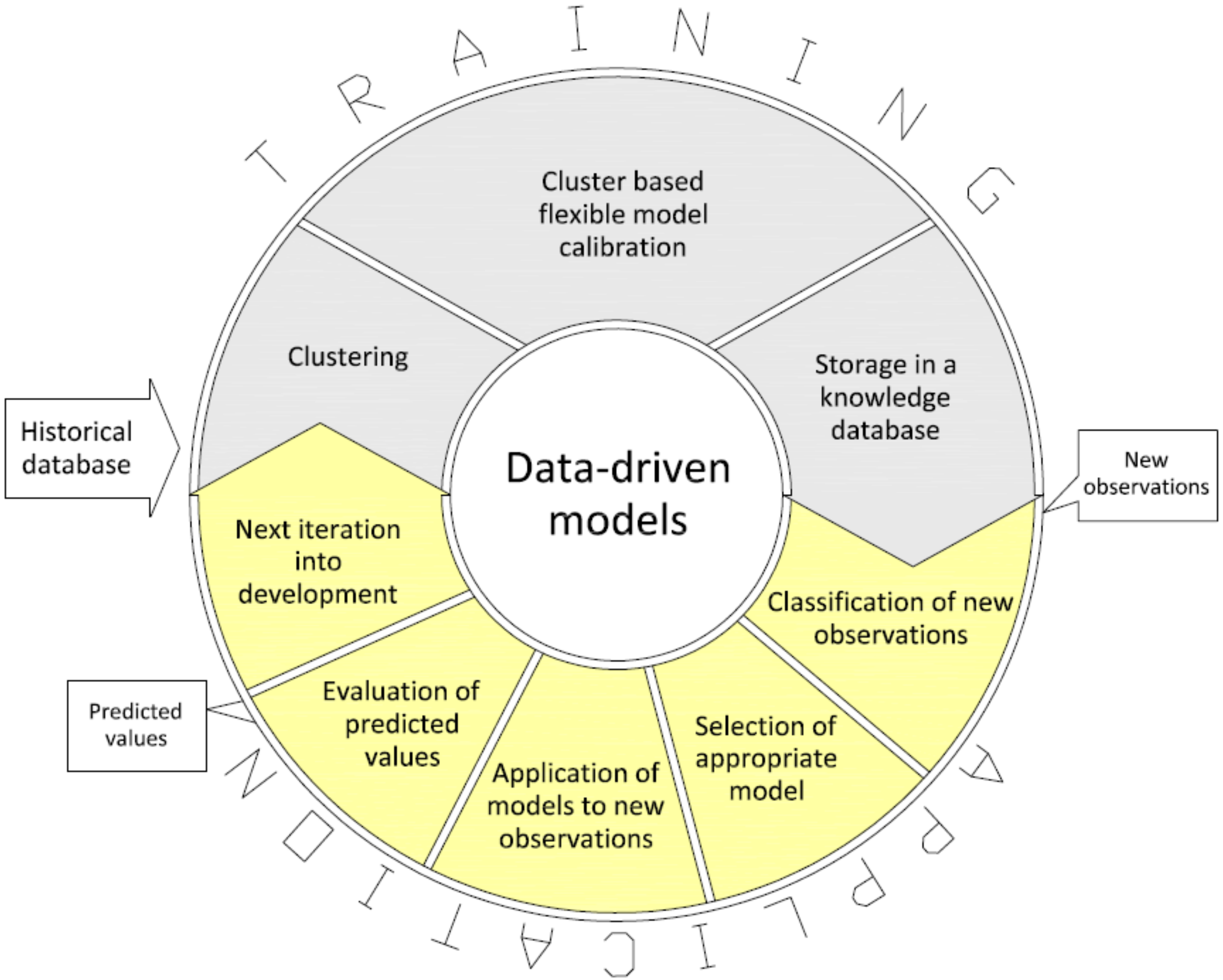


Speed-density relationship

$$u = \begin{cases} u_f & \text{if} \quad k < k_{\min} \\ u_f\left[1 - \left(\left(k - k_{\min}\right)/k_{jam}\right)^{\beta}\right]^{\alpha} & \text{otherwise} \end{cases}$$

Car-following model

$$v_n[t+\tau] = min \begin{cases} v_n[t] + 2.5 \cdot a_n \cdot \tau \cdot (1 - \frac{v_n[t]}{V_n} \cdot \sqrt{(0.025 + \frac{v_n[t]}{V_n})} \\ b_n \cdot \tau + \sqrt{(b_n \cdot \tau)^2 - b_n \cdot [2 \cdot (x_{n-1}[t] - s_{n-1} - x_n[t]) - v_n[t] \cdot \tau - \frac{v_{n-1}^2[t]}{\hat{b}}]} \end{cases}$$


TRAINING
APPLICATION
Data-driven models
Cluster based flexible model calibration
Clustering
Storage in a knowledge database
Historical database
New observations
Next iteration into development
Classification of new observations
Evaluation of predicted values
Predicted values
Application of models to new observations
Selection of appropriate model

Data-driven models

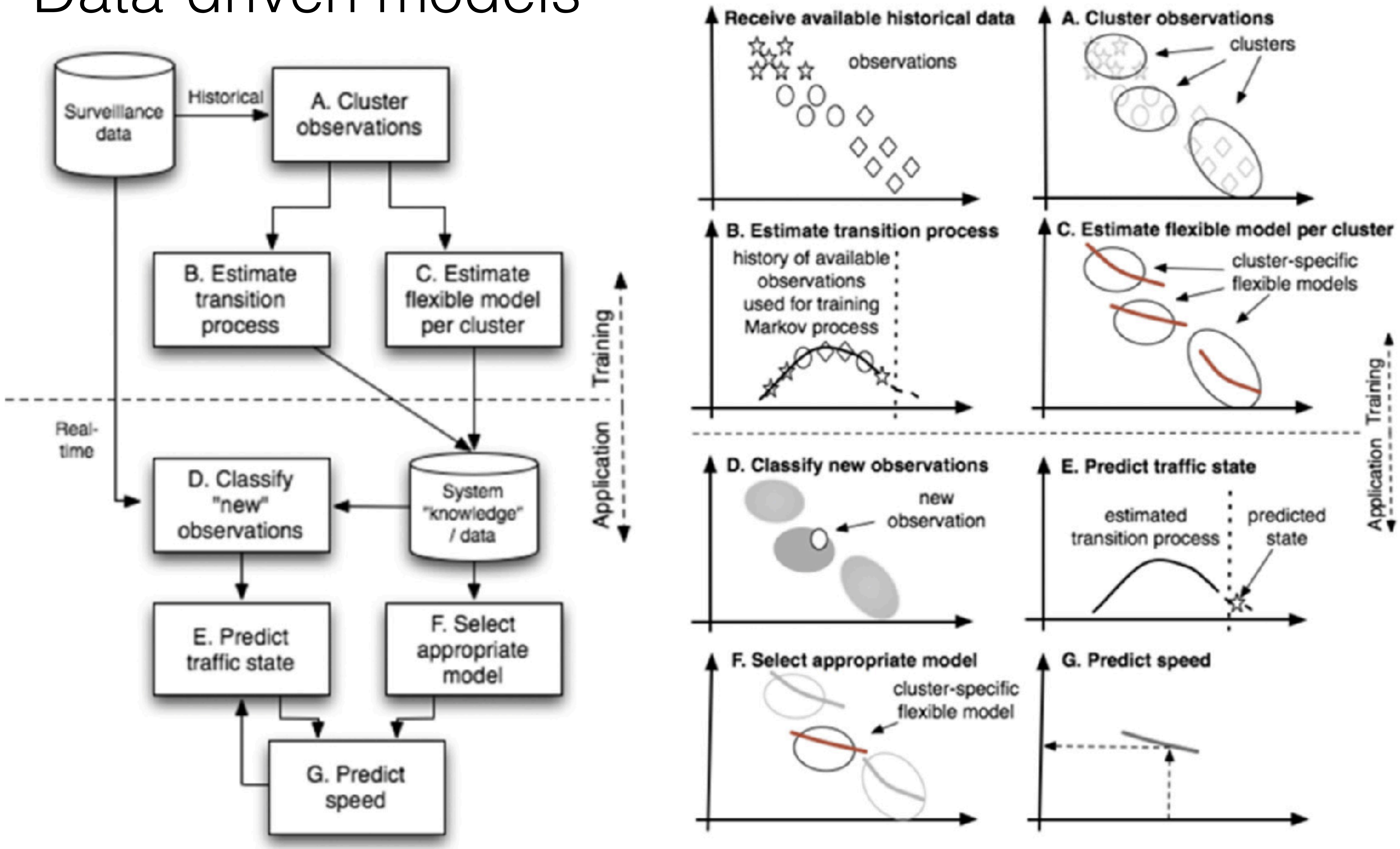
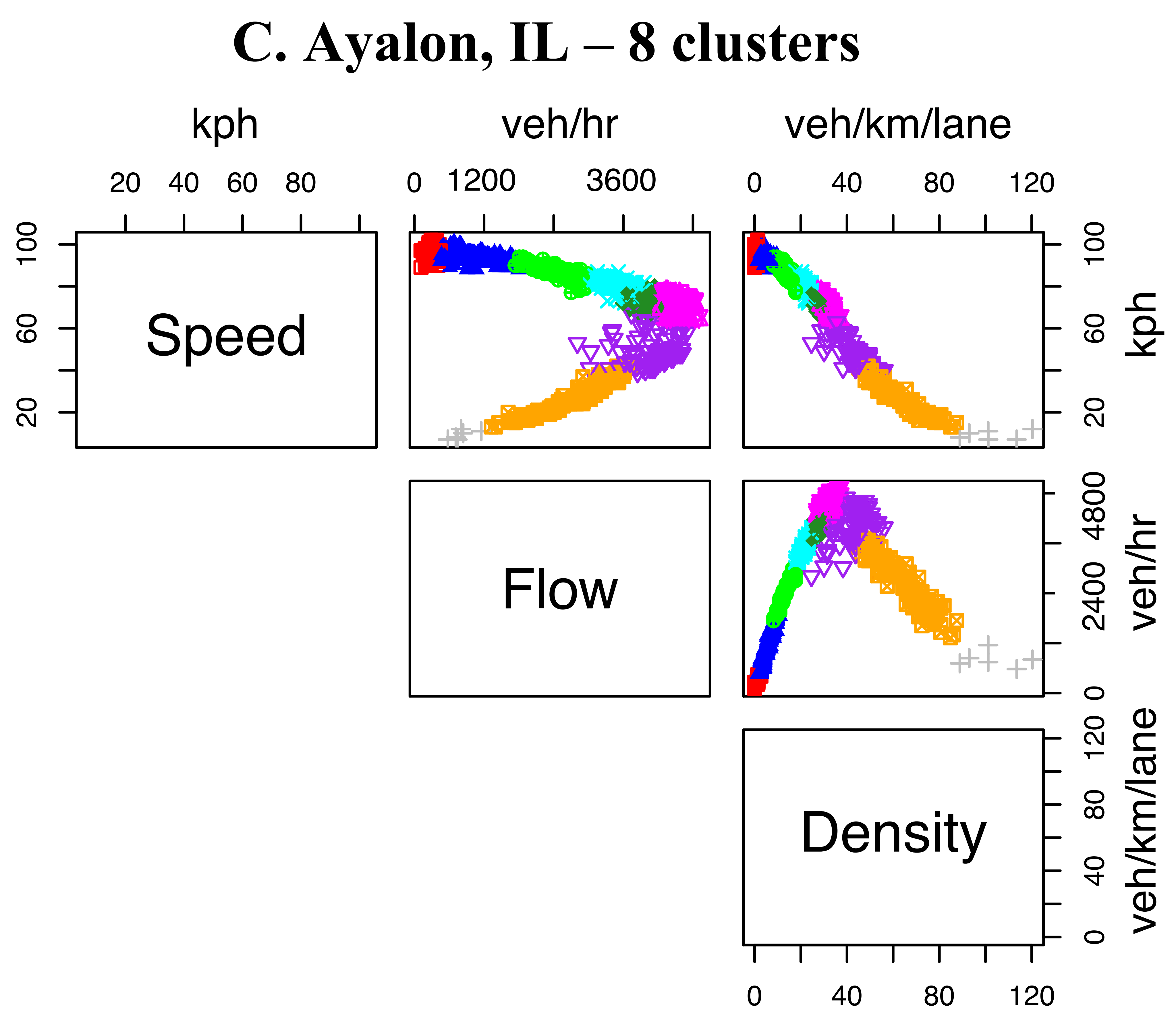
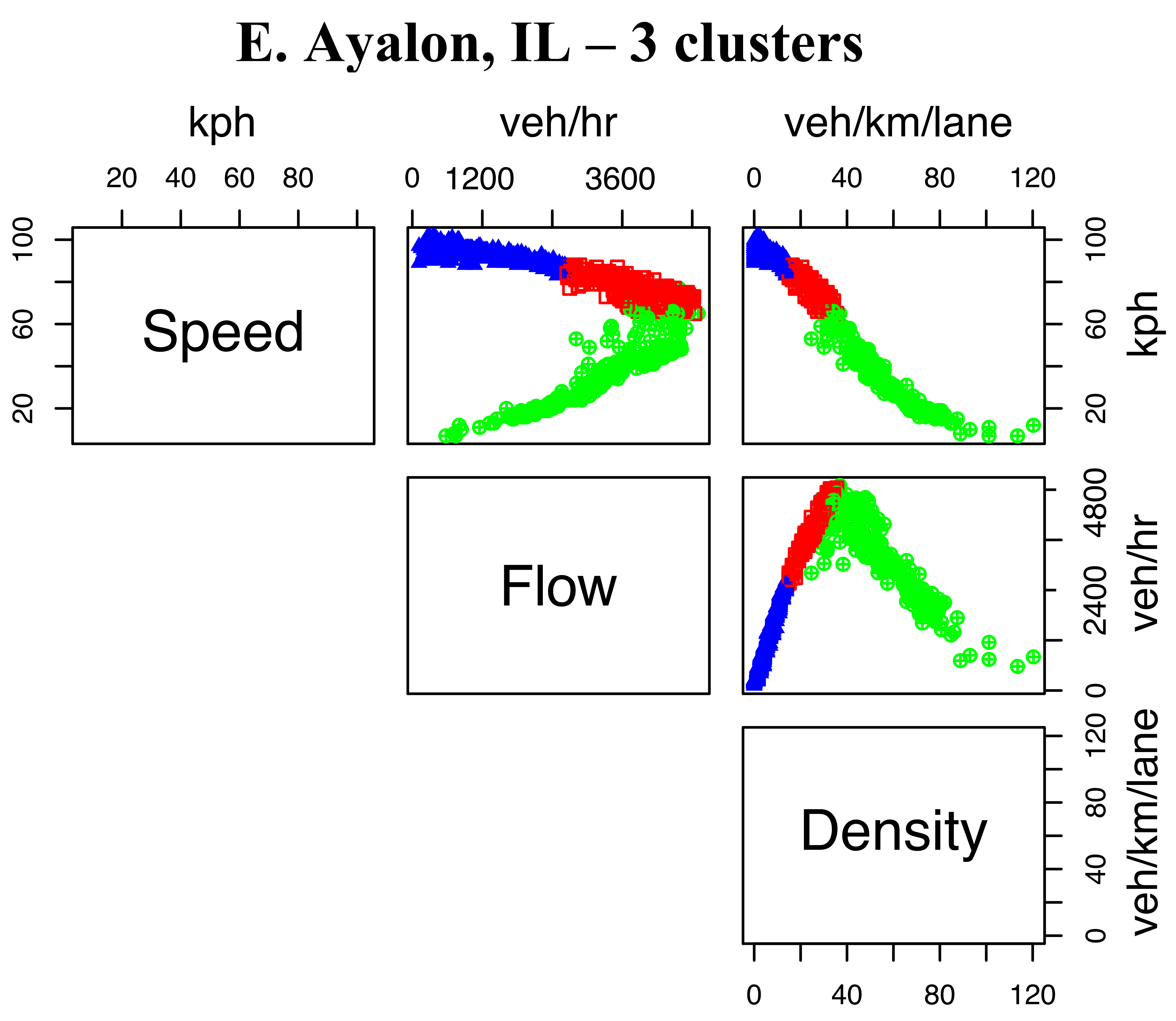


Fig. 1. Overall local traffic state prediction framework.

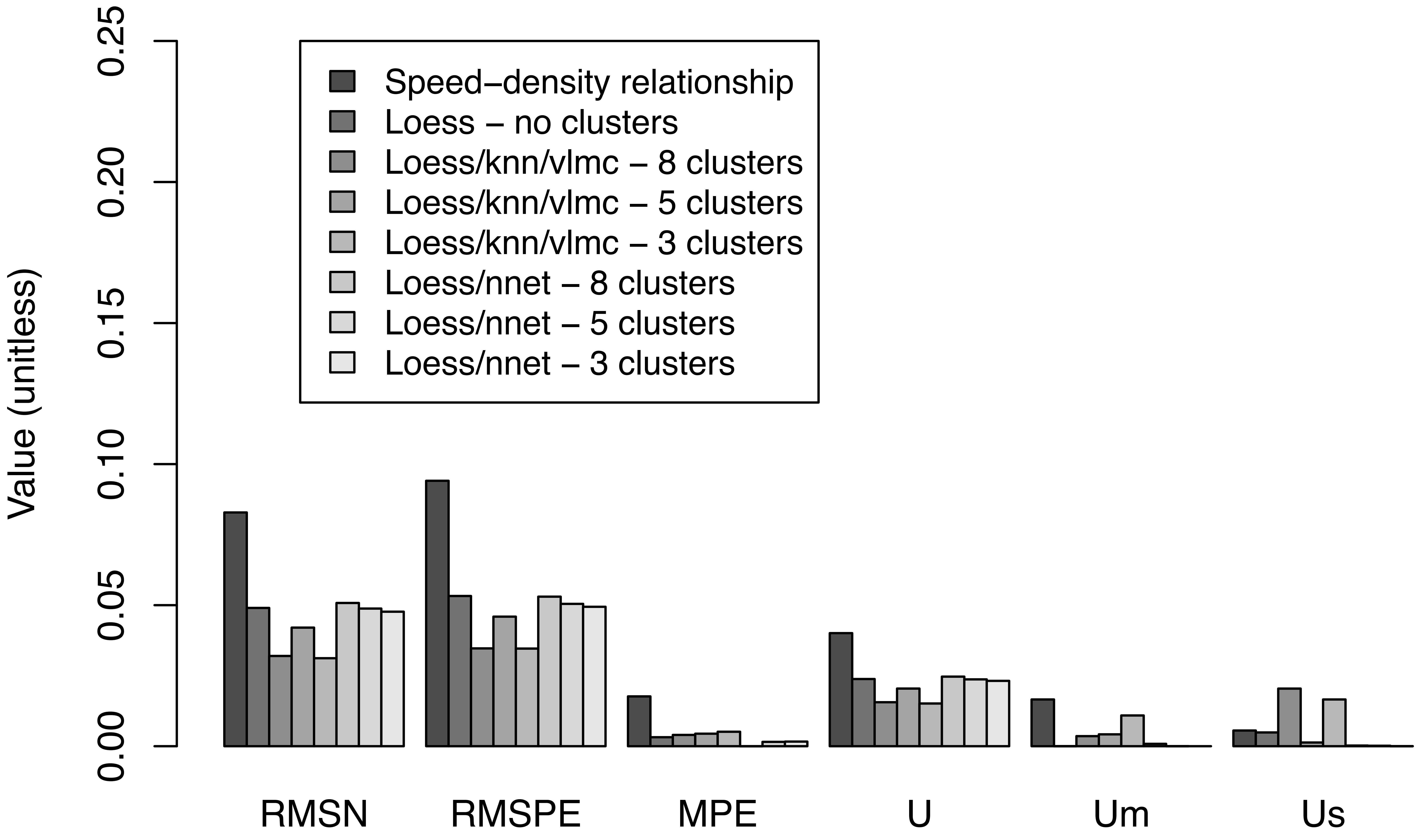
Clustering results (Ayalon motorway, IL)





Antoniou et al., (2013), Transportation Research Part C

Measures of effectiveness – Ayalon, IL



Antoniou et al., (2013), Transportation Research Part C

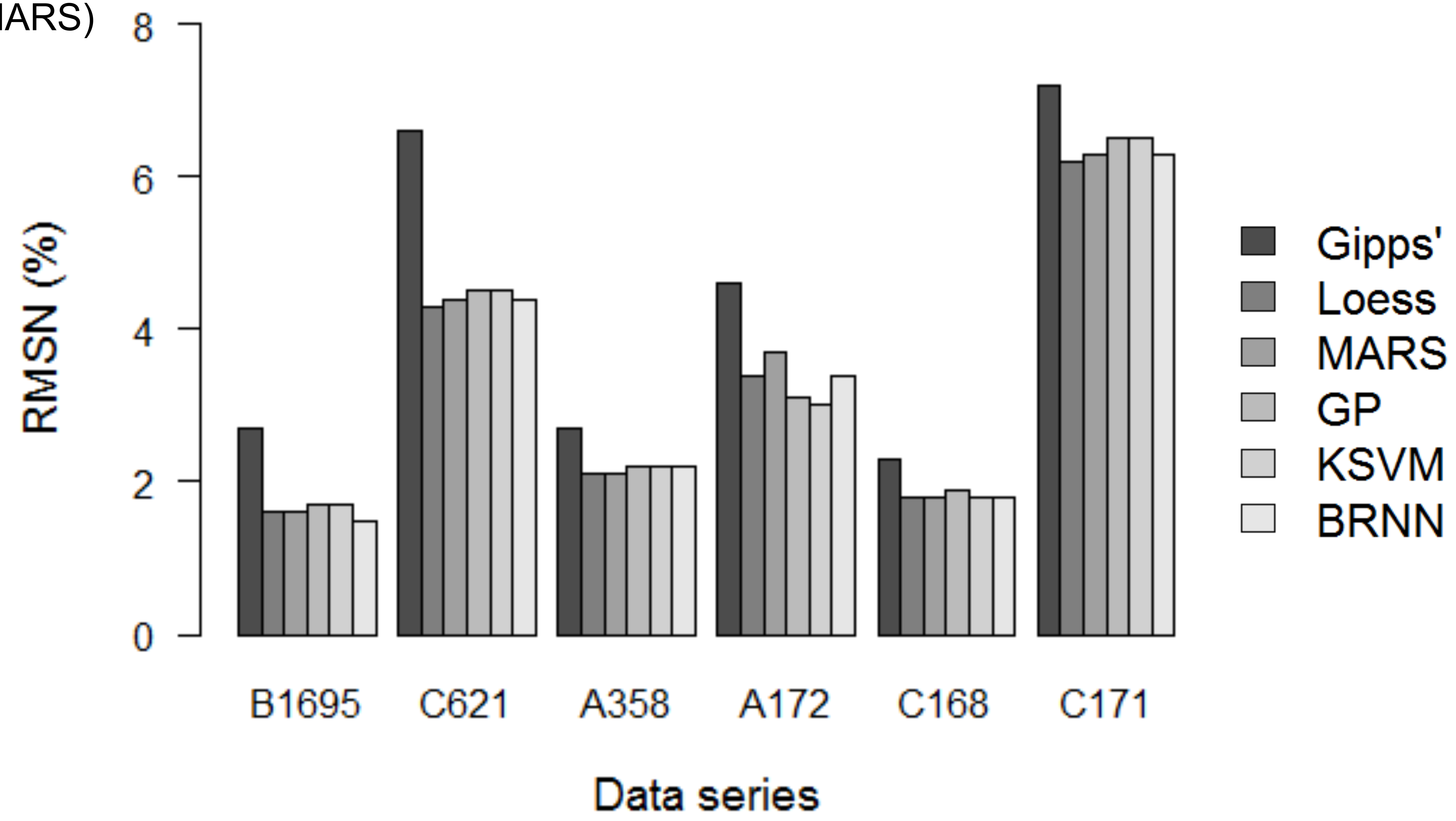
Microscopic case studies

Naples data (Punzo et al., 2005)

NGSIM data

Results - Naples, IT, data

Locally weighted regression (Loess)
Multivariate Adaptive Regression Splines (MARS)
Kernel Support Vector Machines (KSVM)
Gaussian Processes (GP)
Bayesian Regularized Neural
Networks (BRNN)



Lehrstuhl für Vernetzte Verkehrssysteme
Fakultät für Bau, Geo und Umwelt
Technische Universität München

Mixed traffic conditions

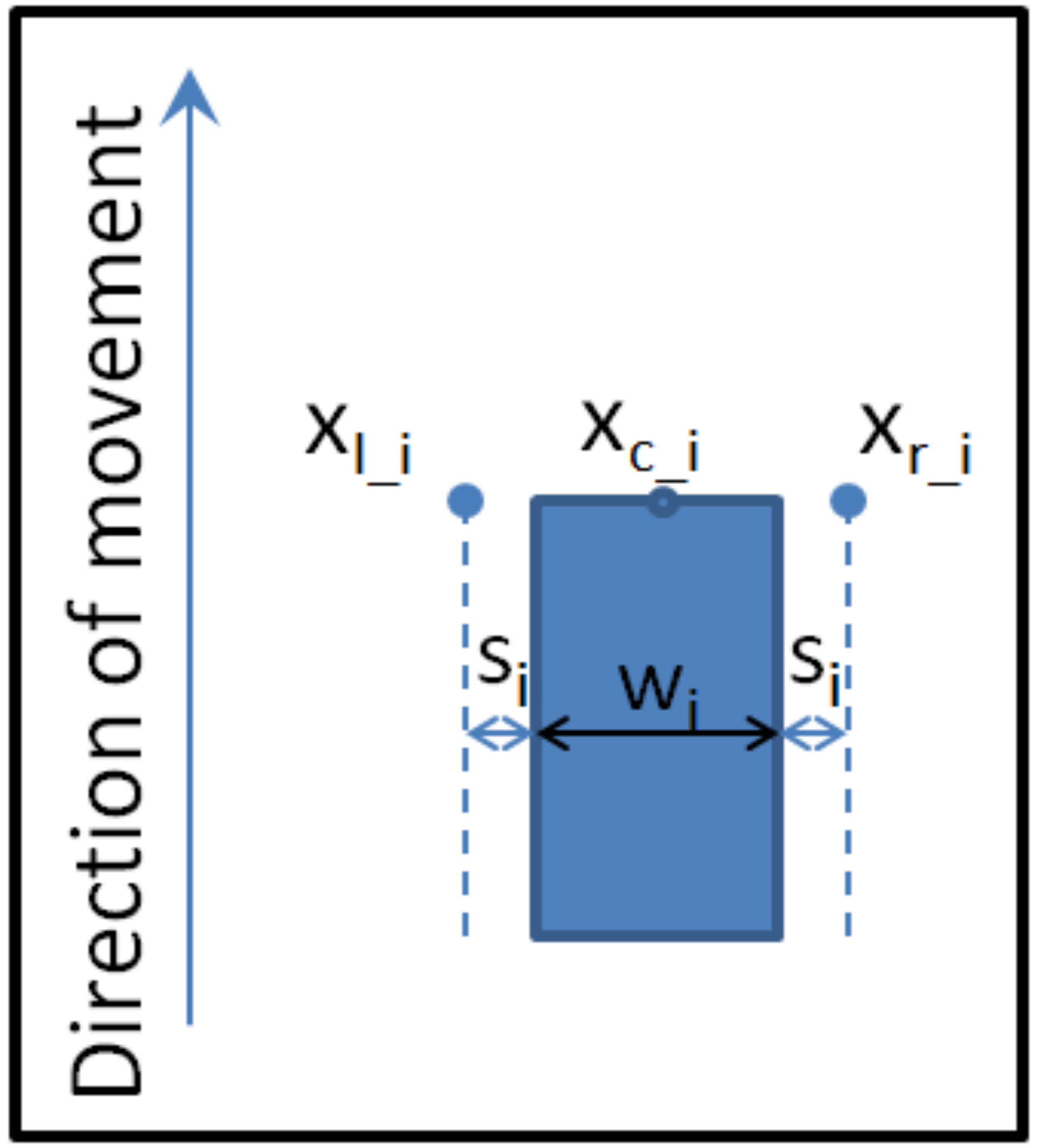
Weak lane discipline
Multiple vehicle types

Video data were collected on February 13, 2014, on a six-lane separated urban arterial road at the Maraimalai Adigalar Bridge in Saidapet, Chennai, India (Kanagaraj et al., 2015). The trajectory data are shared publicly at the address: http://toledo.net.technion.ac.il/downloads/.

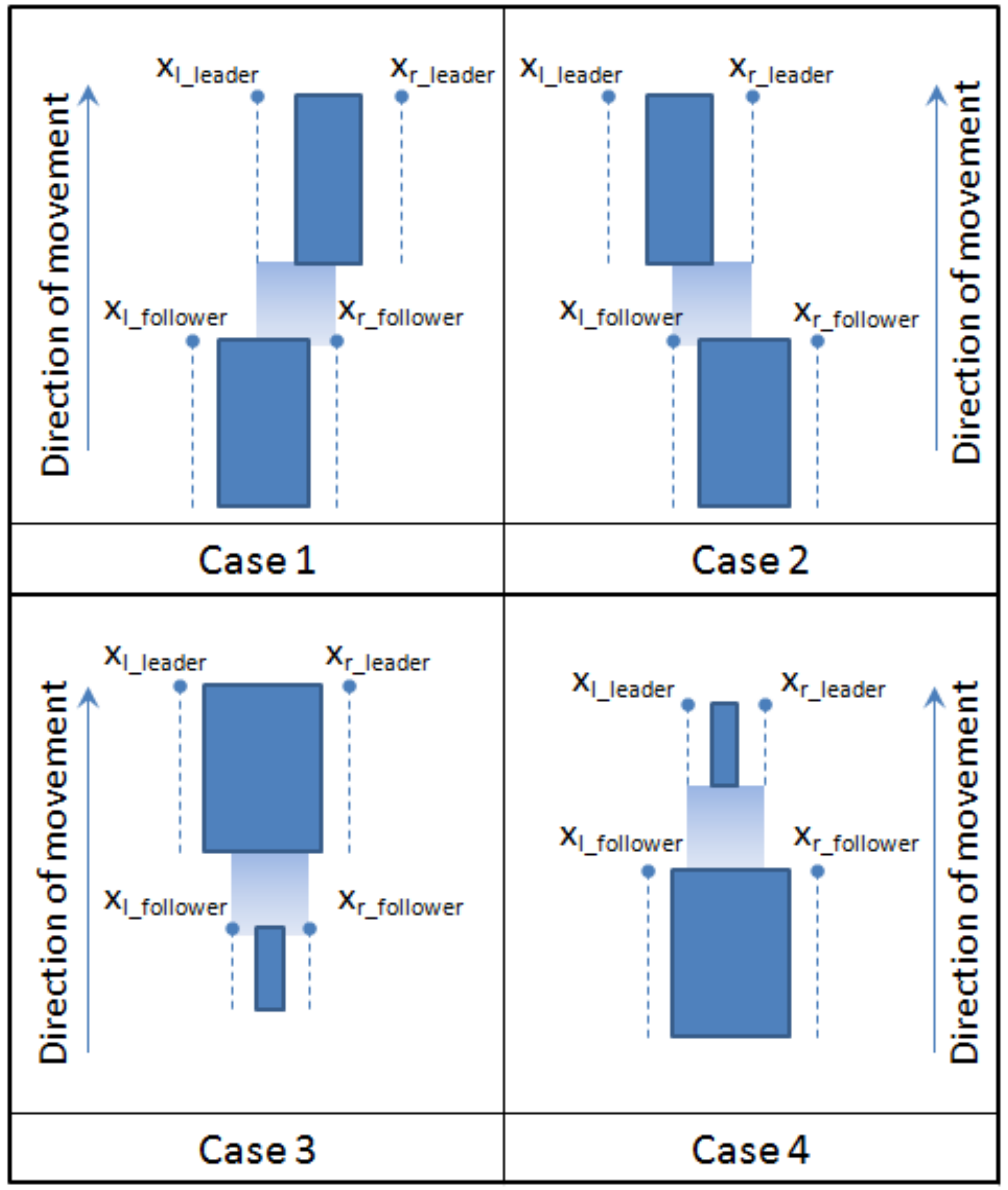
•Data for model calibration: data245 (data collected in the period 2:45-3:00 PM)
•Data for model validation: data300 (data collected in the period 3:00-3:15 PM)

Lehrstuhl für Vernetzte Verkehrssysteme
Fakultät für Bau, Geo und Umwelt
Technische Universität München

Mixed traffic conditions - Modeling

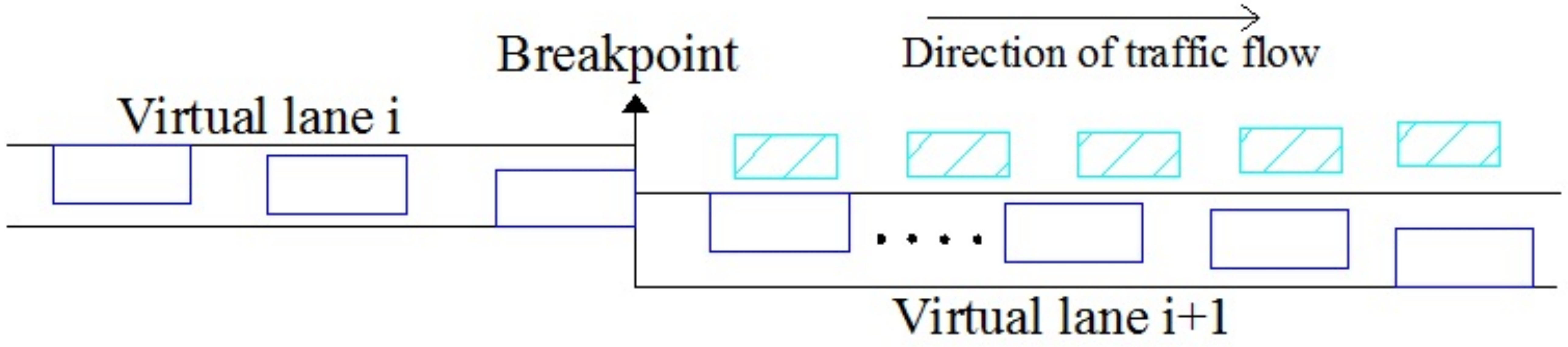


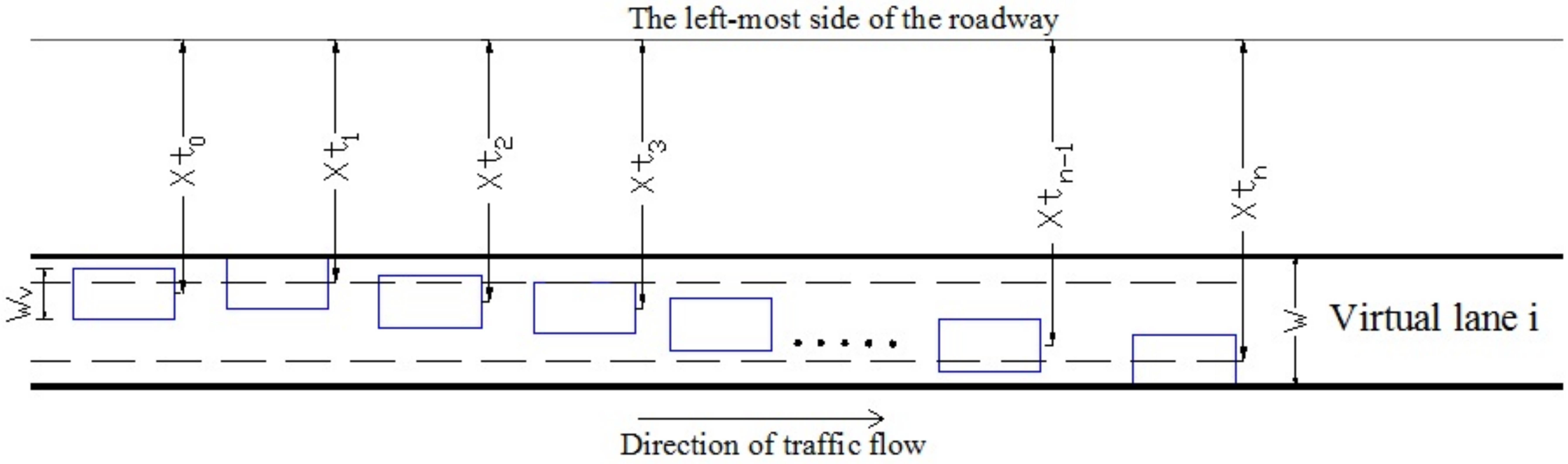
Estimation of lateral bounds of each vehicle



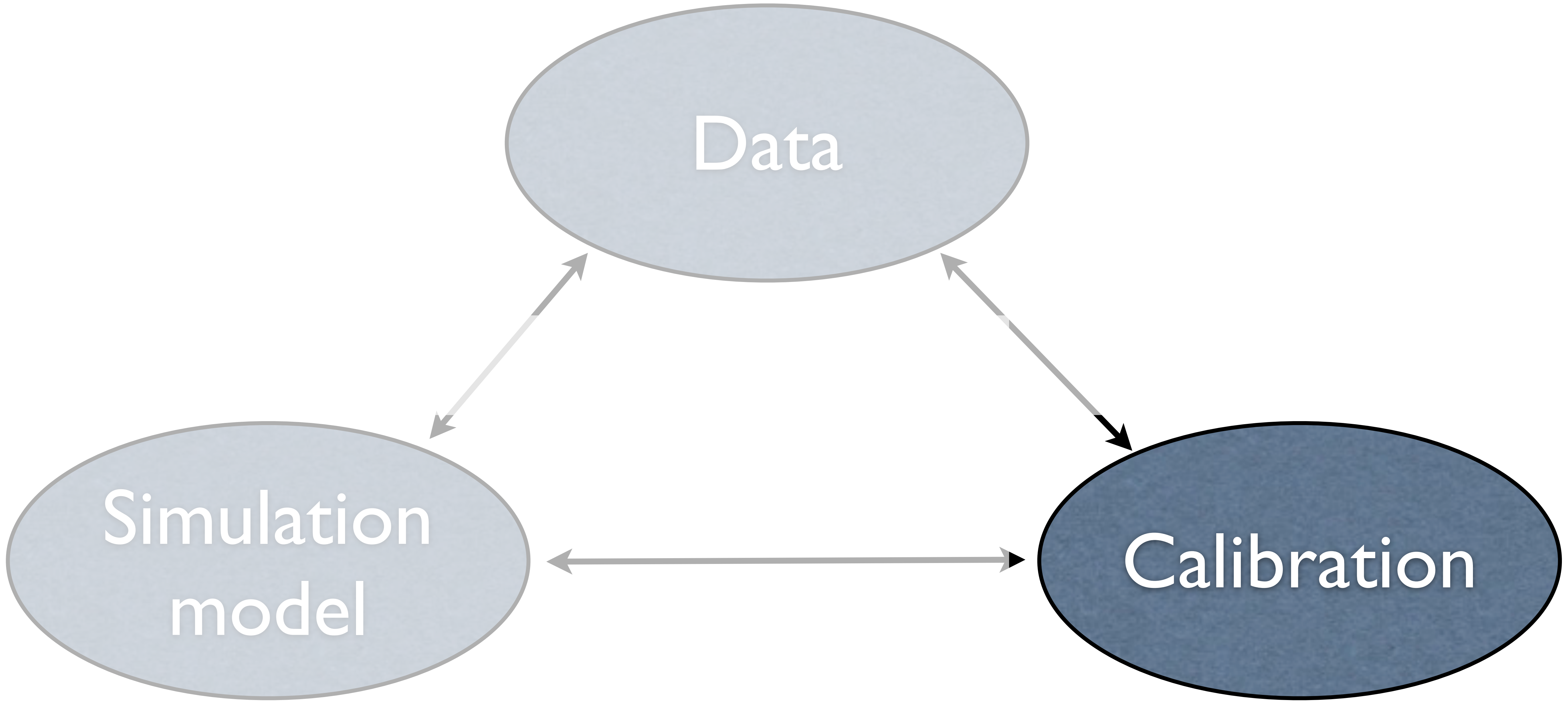
Identification of leader-follower pair

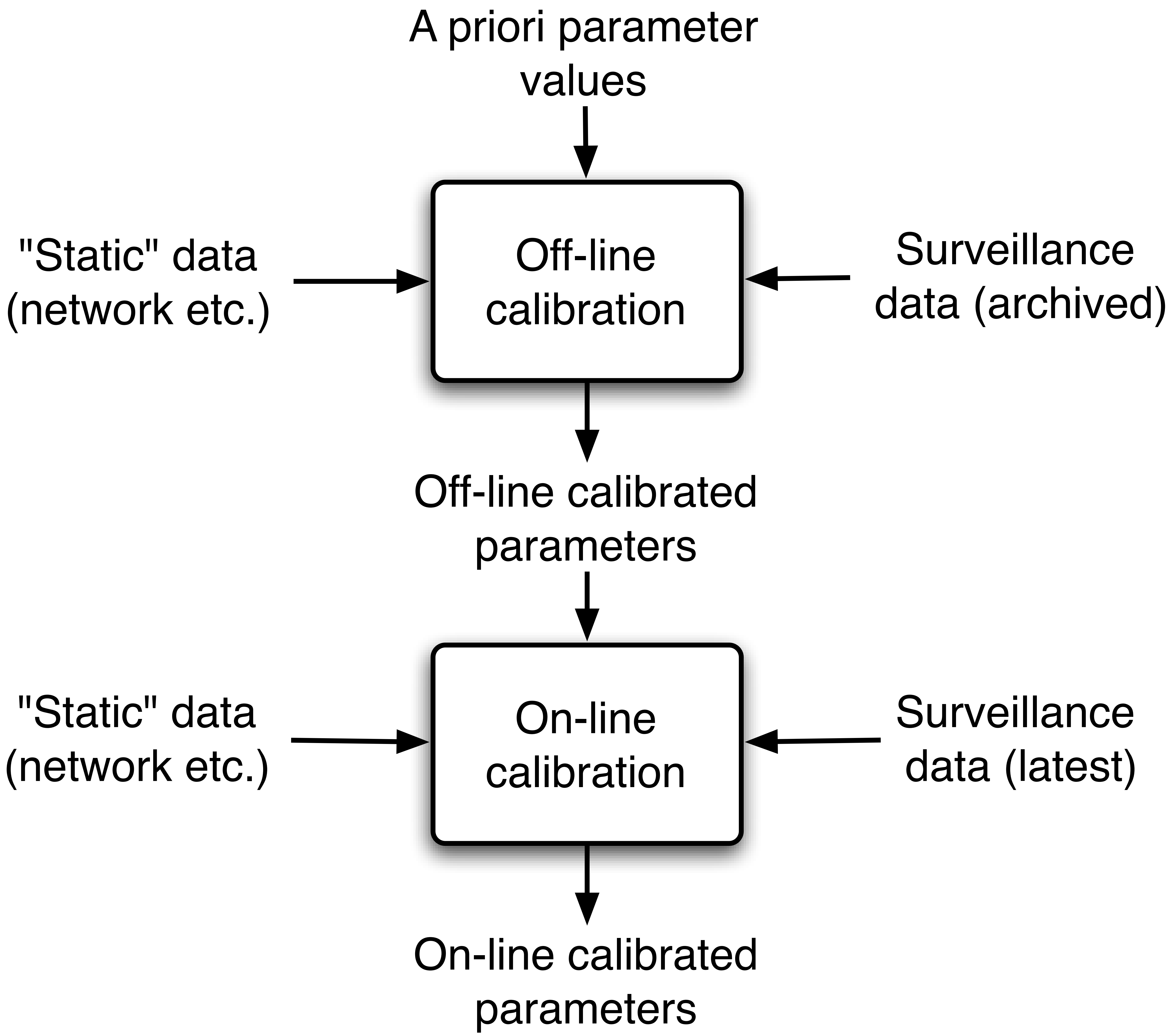
Mixed traffic conditions - Modeling (cont'd)



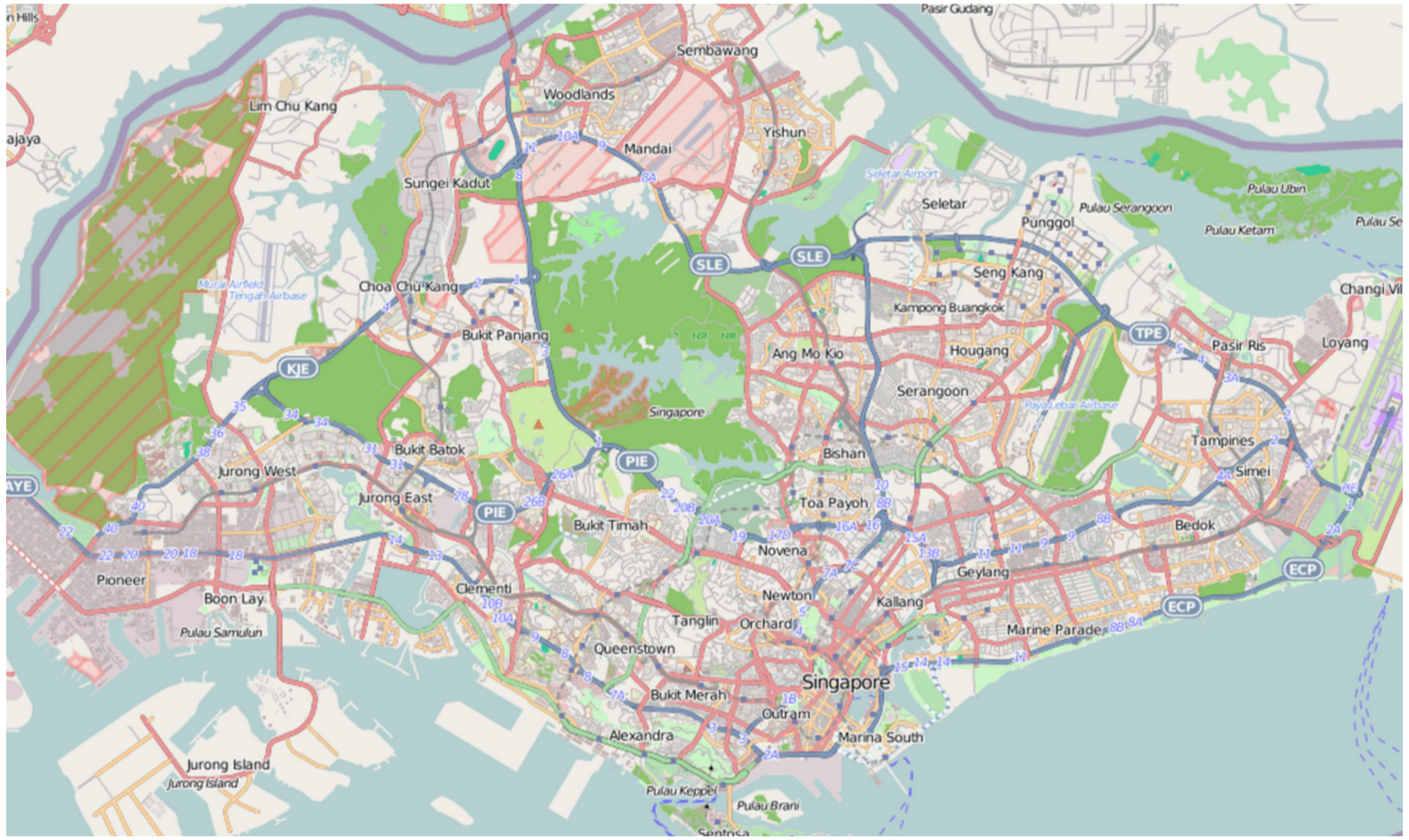


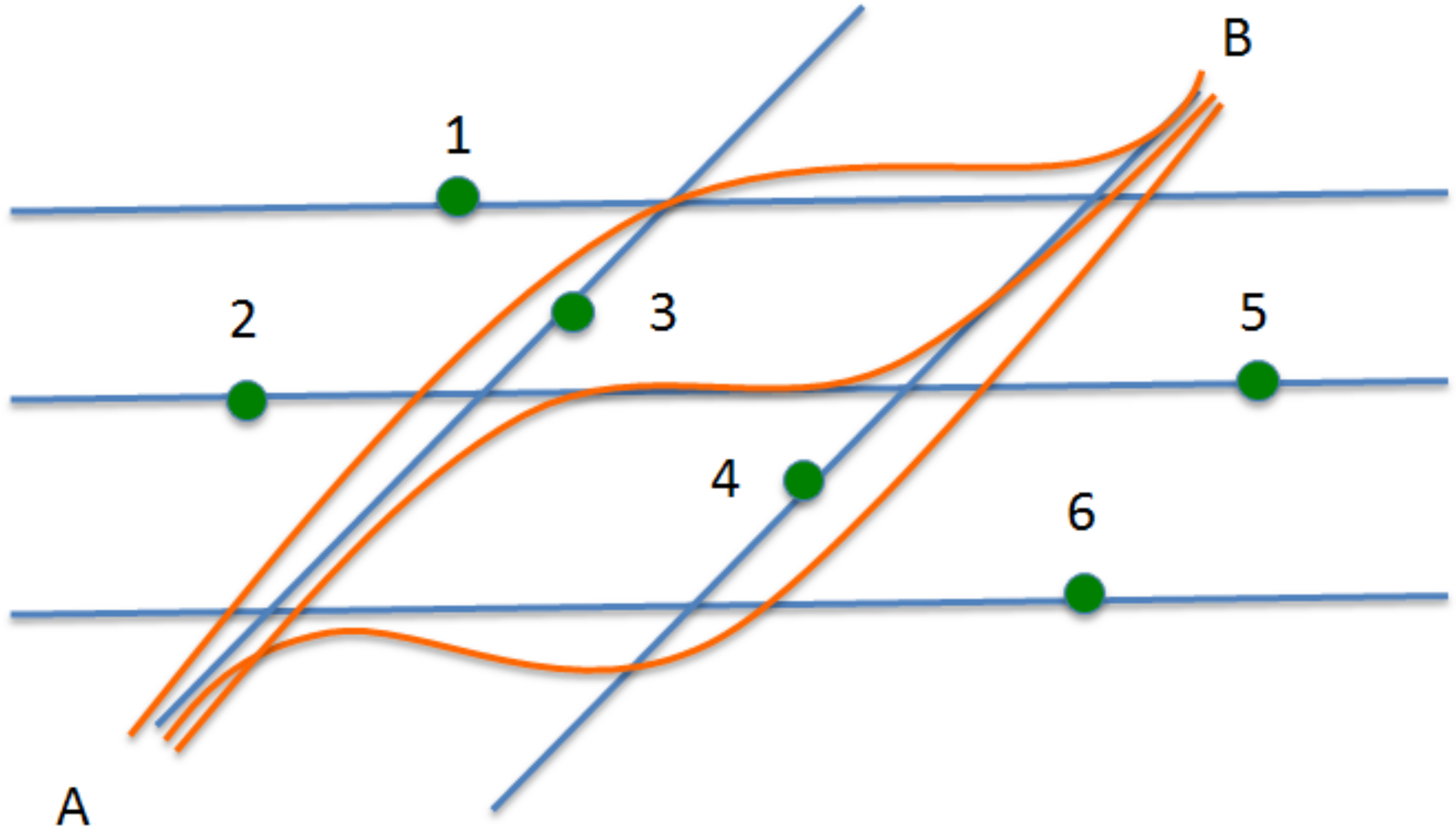
$$W=\max(x_{t0}, x_{t1}, \ldots, x_{tn}) - \min(x_{t0}, x_{t1}, \ldots, x_{tn}) + w_v$$


Data
Simulation model
Calibration


A priori parameter values
"Static" data (network etc.)
Off-line calibration
Surveillance data (archived)
Off-line calibrated parameters
"Static" data (network etc.)
On-line calibration
Surveillance data (latest)
On-line calibrated parameters

Dealing with really large-scale problems — W-SPSA



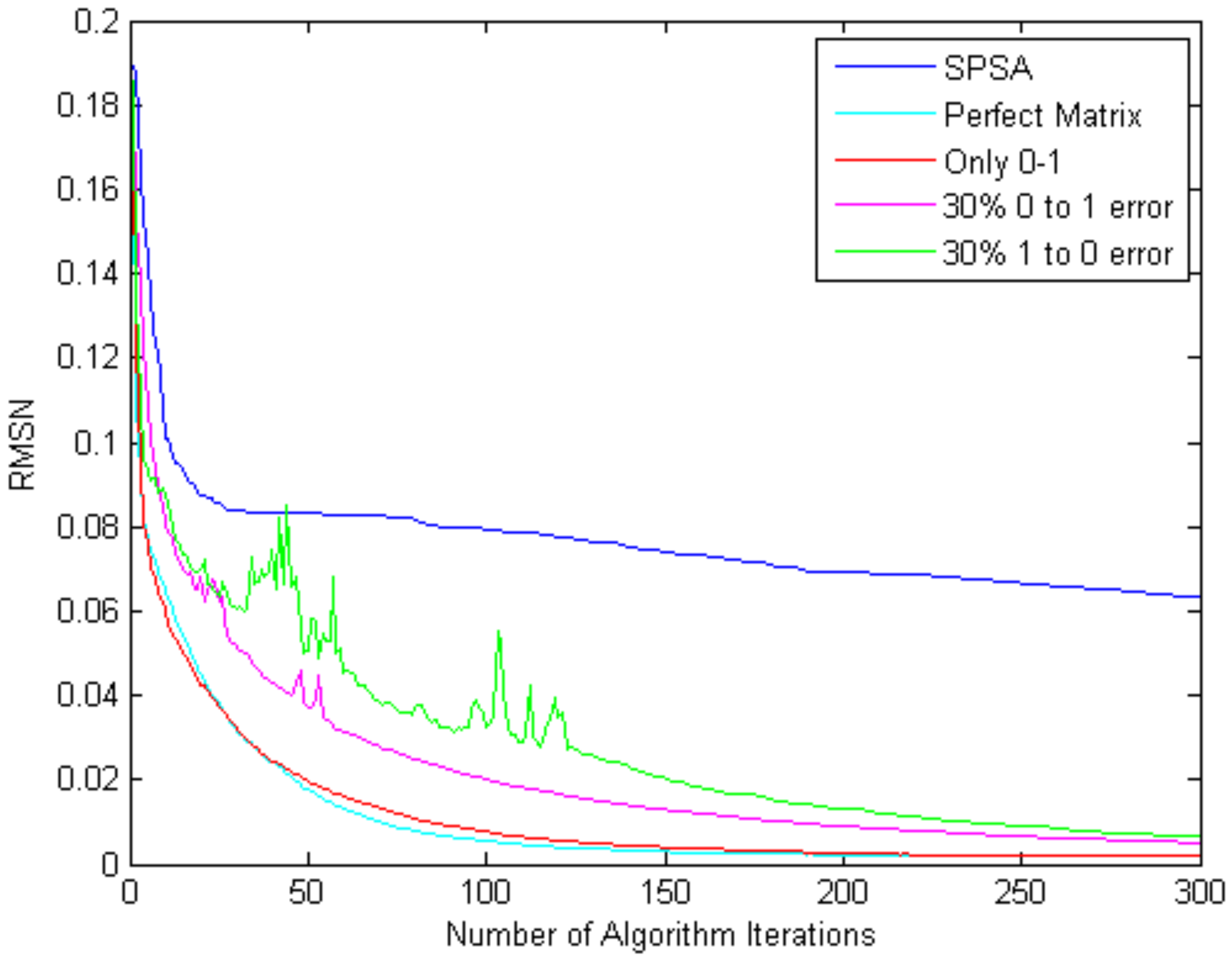


Lu, L., Y. Xu, C. Antoniou and M. Ben-Akiva (2015), W-SPSA: An Enhanced SPSA Algorithm for the Calibration of Dynamic Traffic Assignment Models, Transportation Research: Part C, 51, pp. 149-166

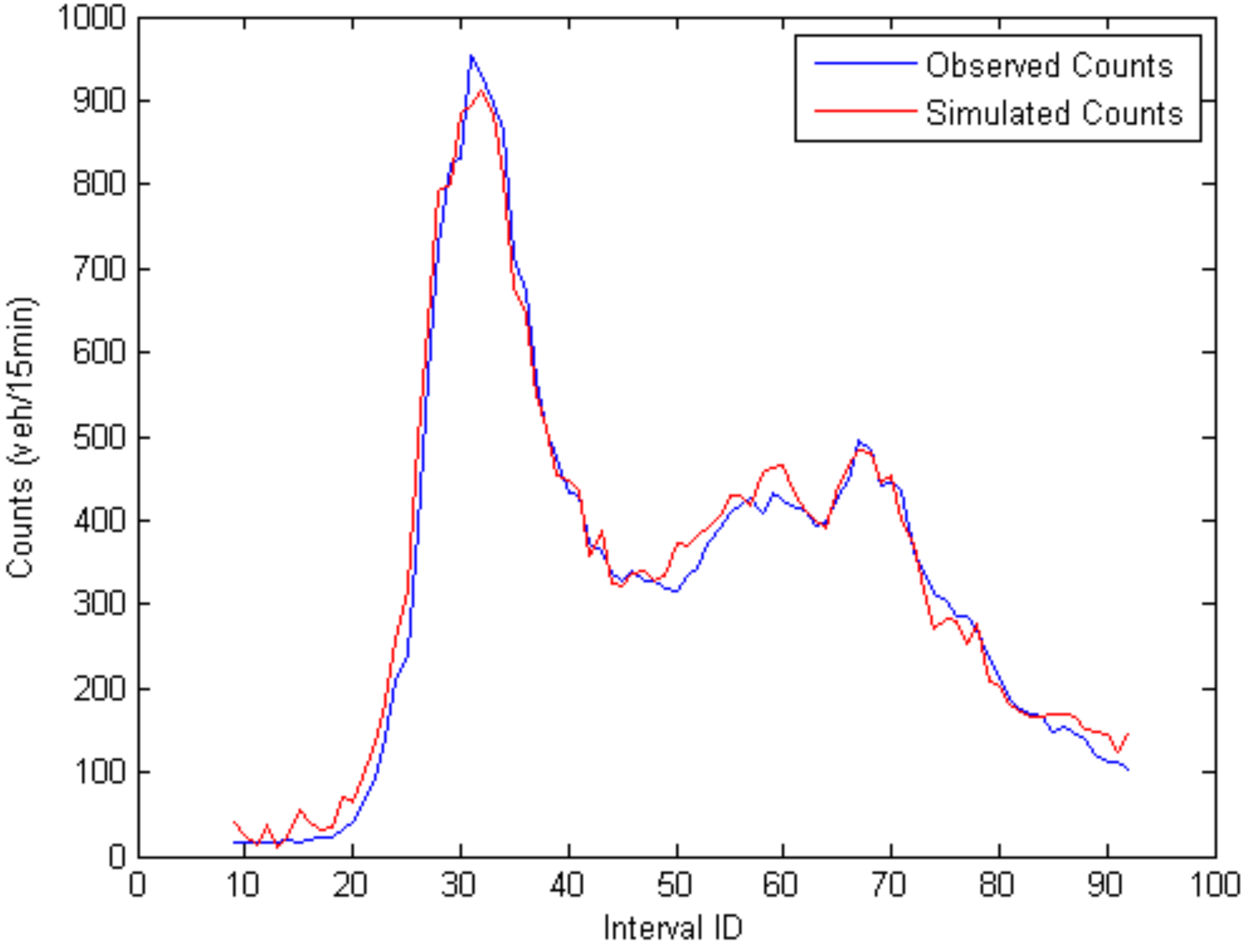
Antoniou, C., C. L. Azevedo, L. Lu, F. Pereira and M. Ben-Akiva (2015). W-SPSA in practice: Approximation of weight matrices and calibration of traffic simulation models. Transportation Research Part C, Vol. 59, pp. 129-146.

W-SPSA results

Synthetic experiment



Singapore Expressway Network

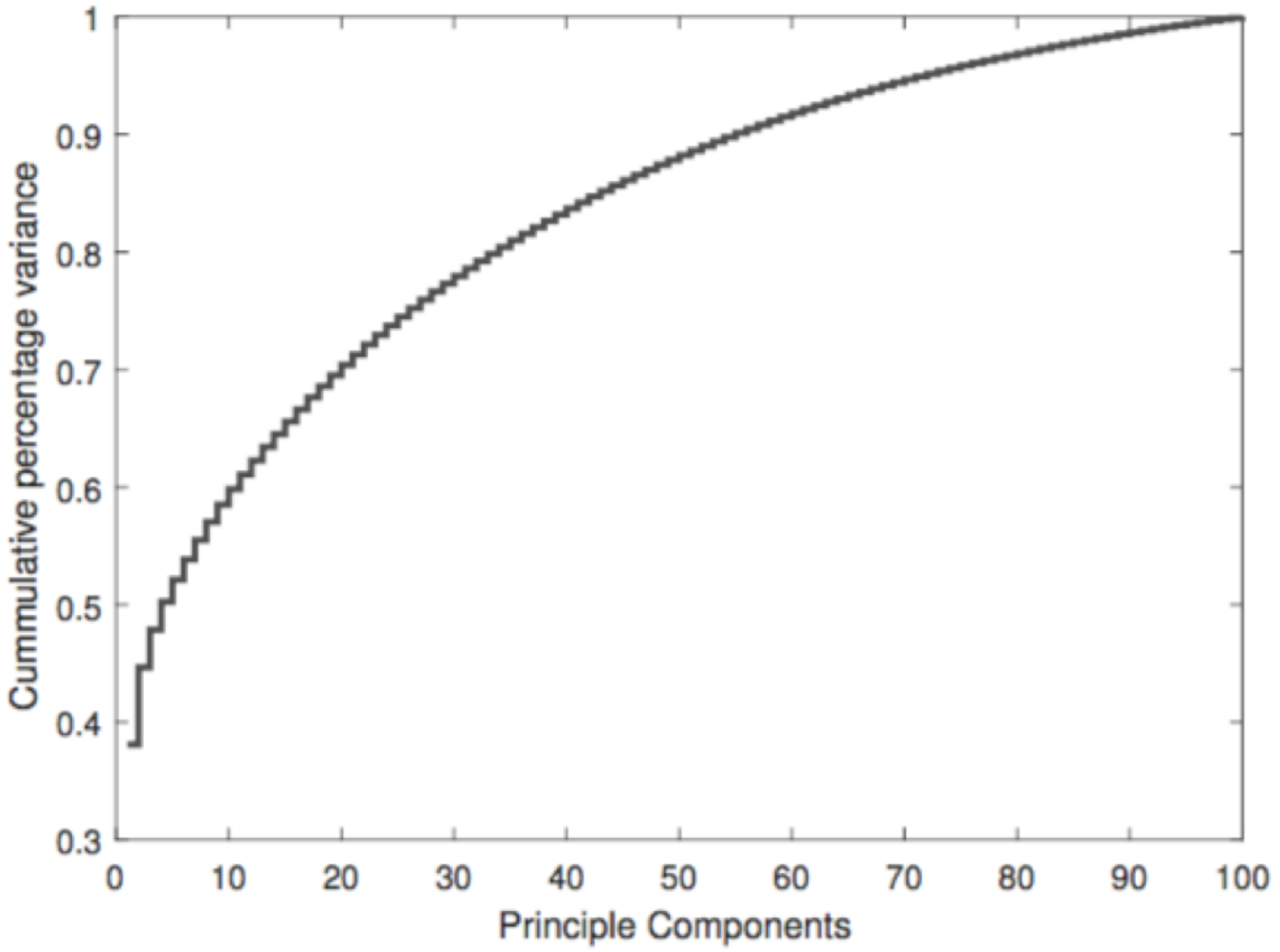


Lu, L., Y. Xu, C. Antoniou and M. Ben-Akiva (2015), W-SPSA: An Enhanced SPSA Algorithm for the Calibration of Dynamic Traffic Assignment Models, Transportation Research: Part C, 51, pp. 149-166

Dimensionality reduction

Principal Component (PC) Analysis

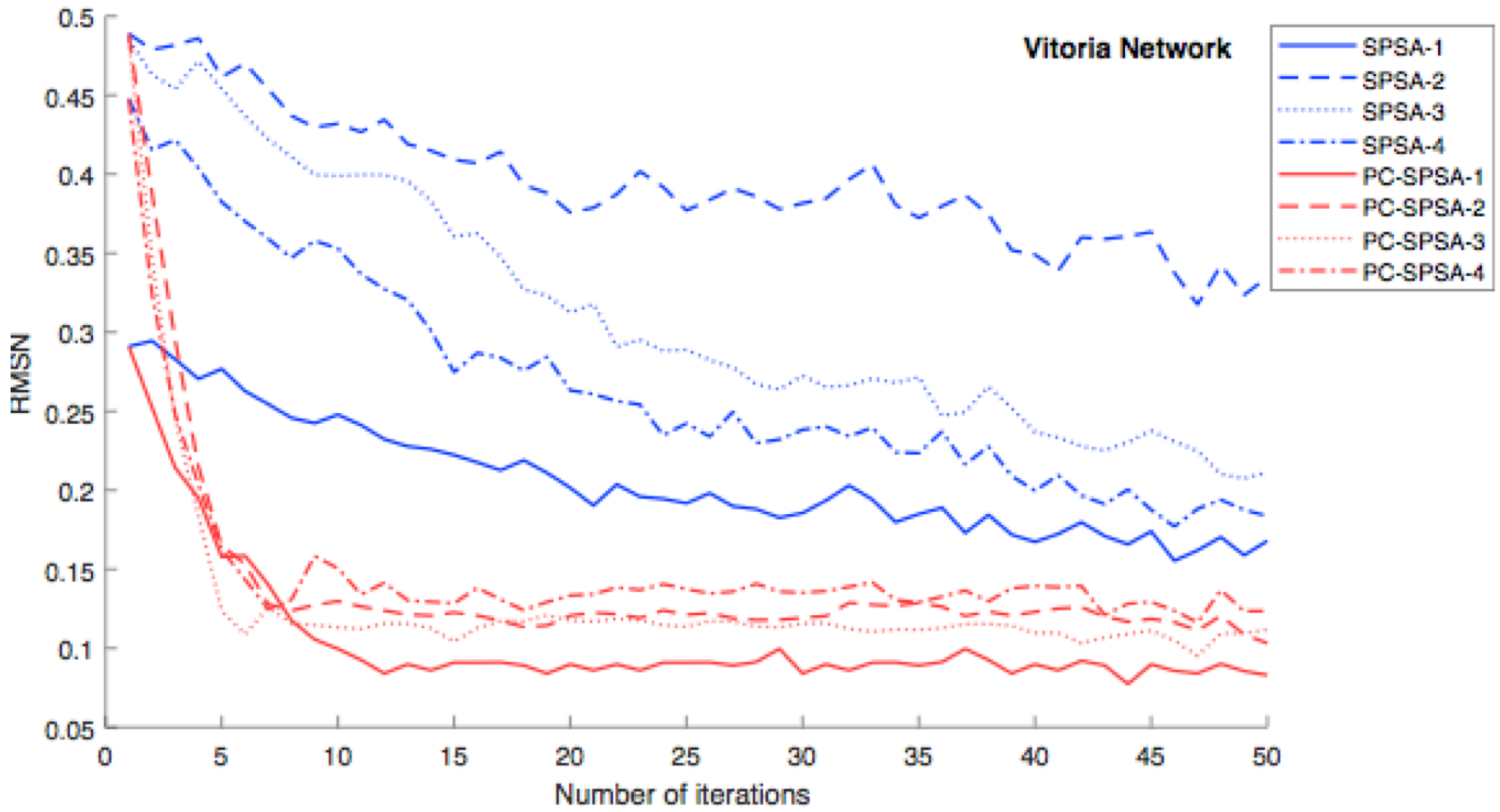
Has led to many algorithms, such as
PC-GLS (Prakash et al., 2017)
PC-EKF (Prakash et al., 2018)
PC-SPSA (Qurashi et al., 2018)



Vitoria network: 80 PCs capture >95% of variance

PC-SPSA Case Study — Vitoria, ES

57 zones => 57x56=3192 OD pairs per 15 min interval



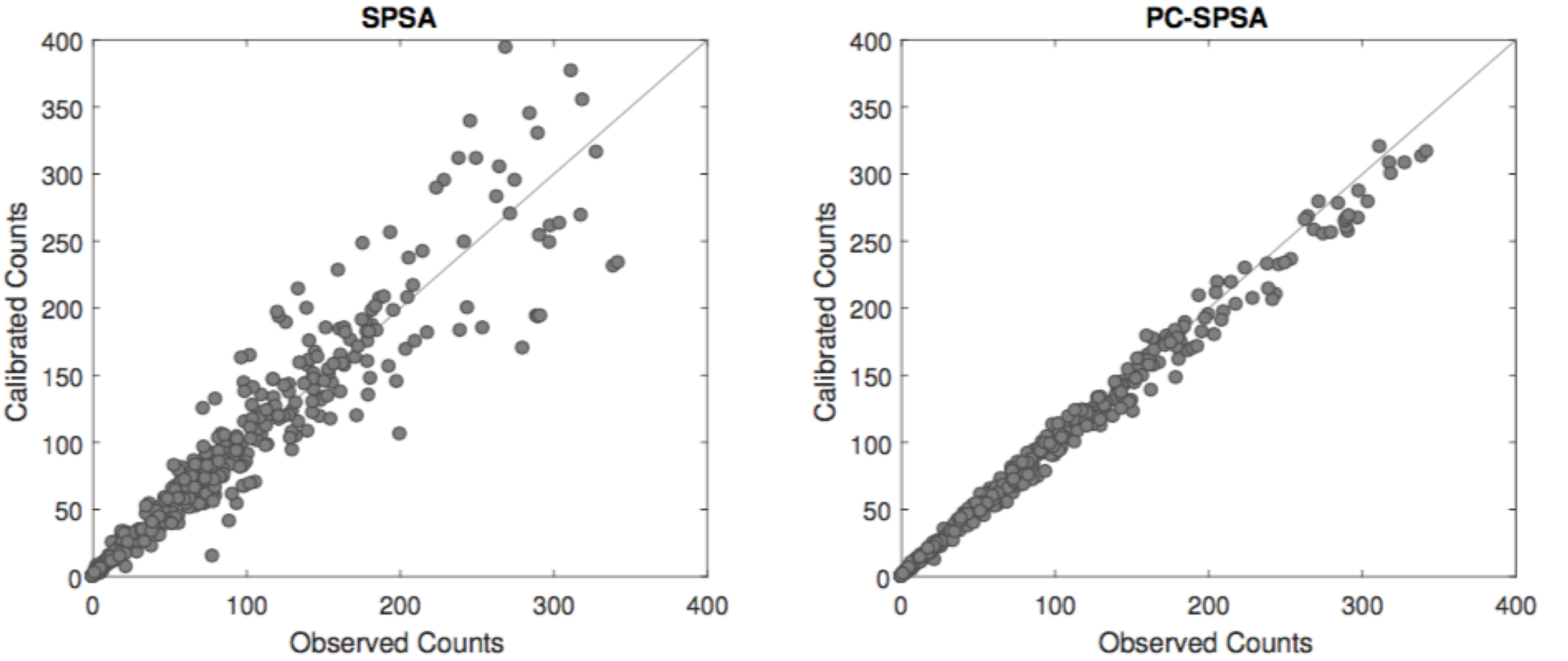
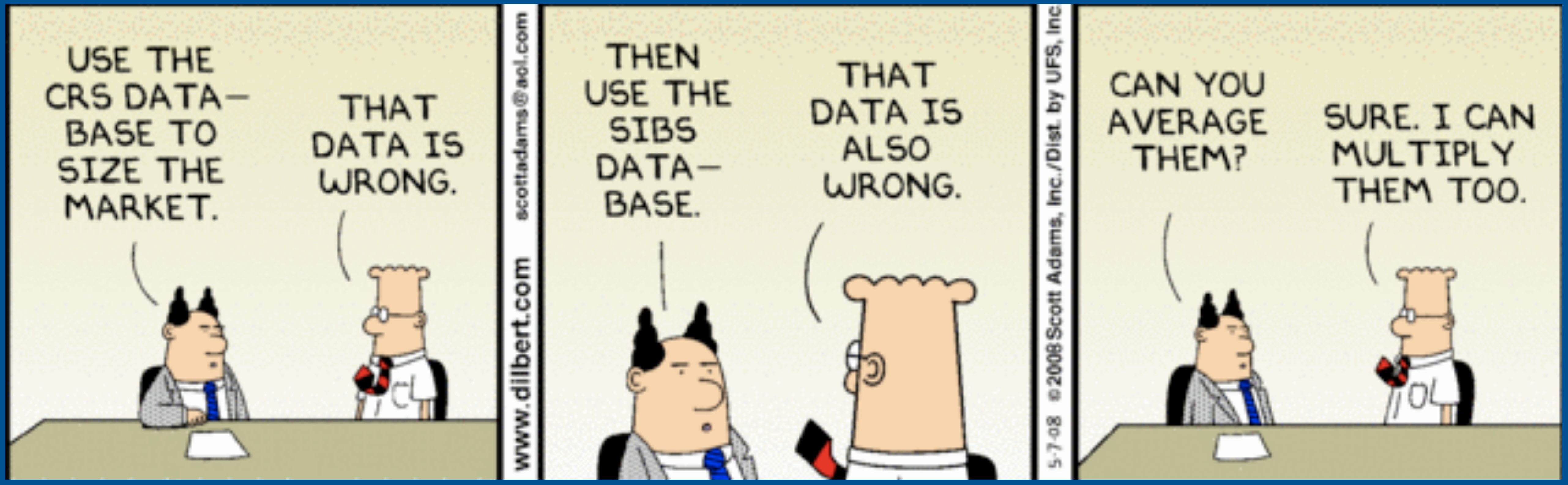
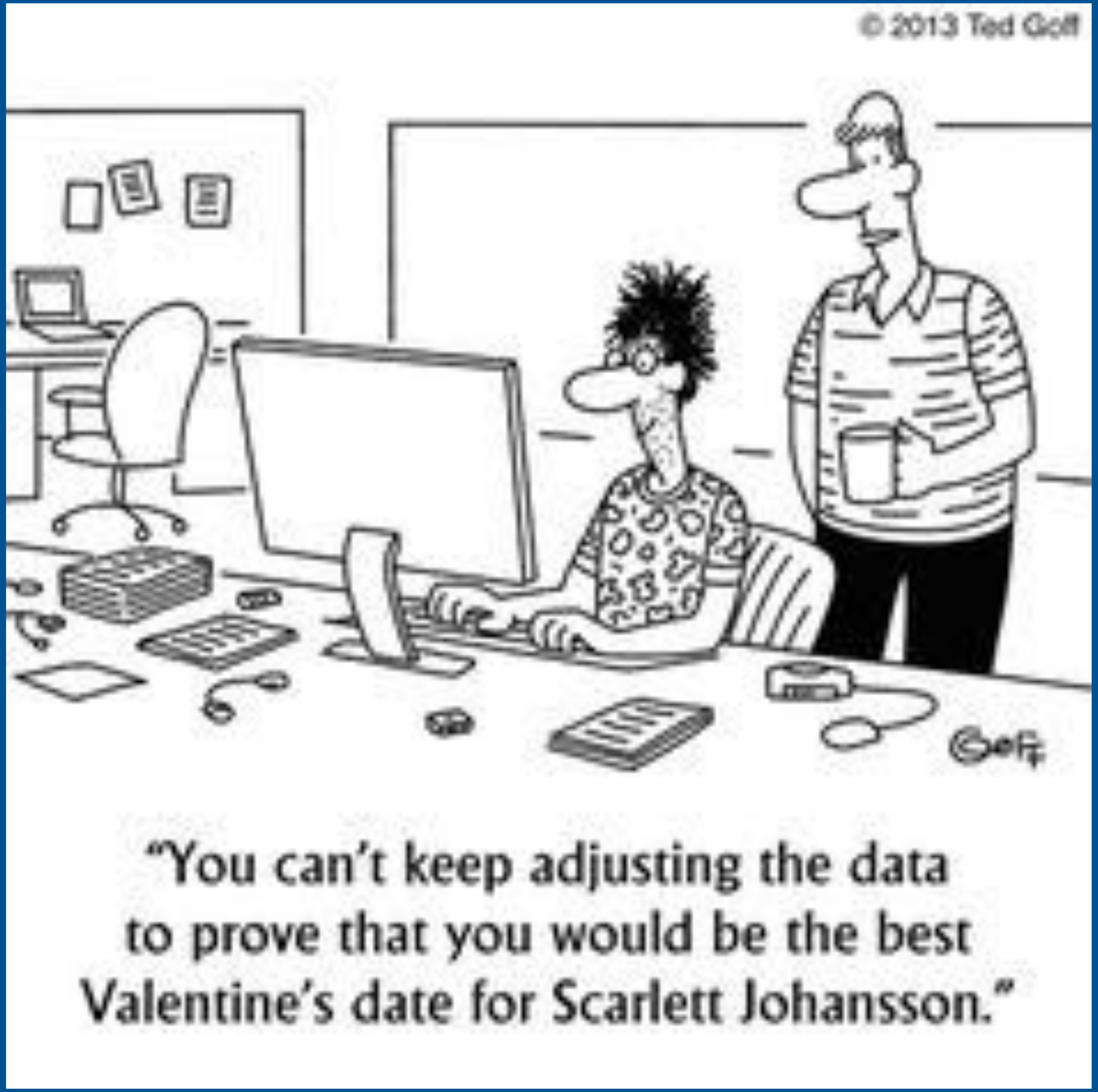


Figure 4.16: Calibrated Counts by SPSA and PC-SPSA

Pitfalls

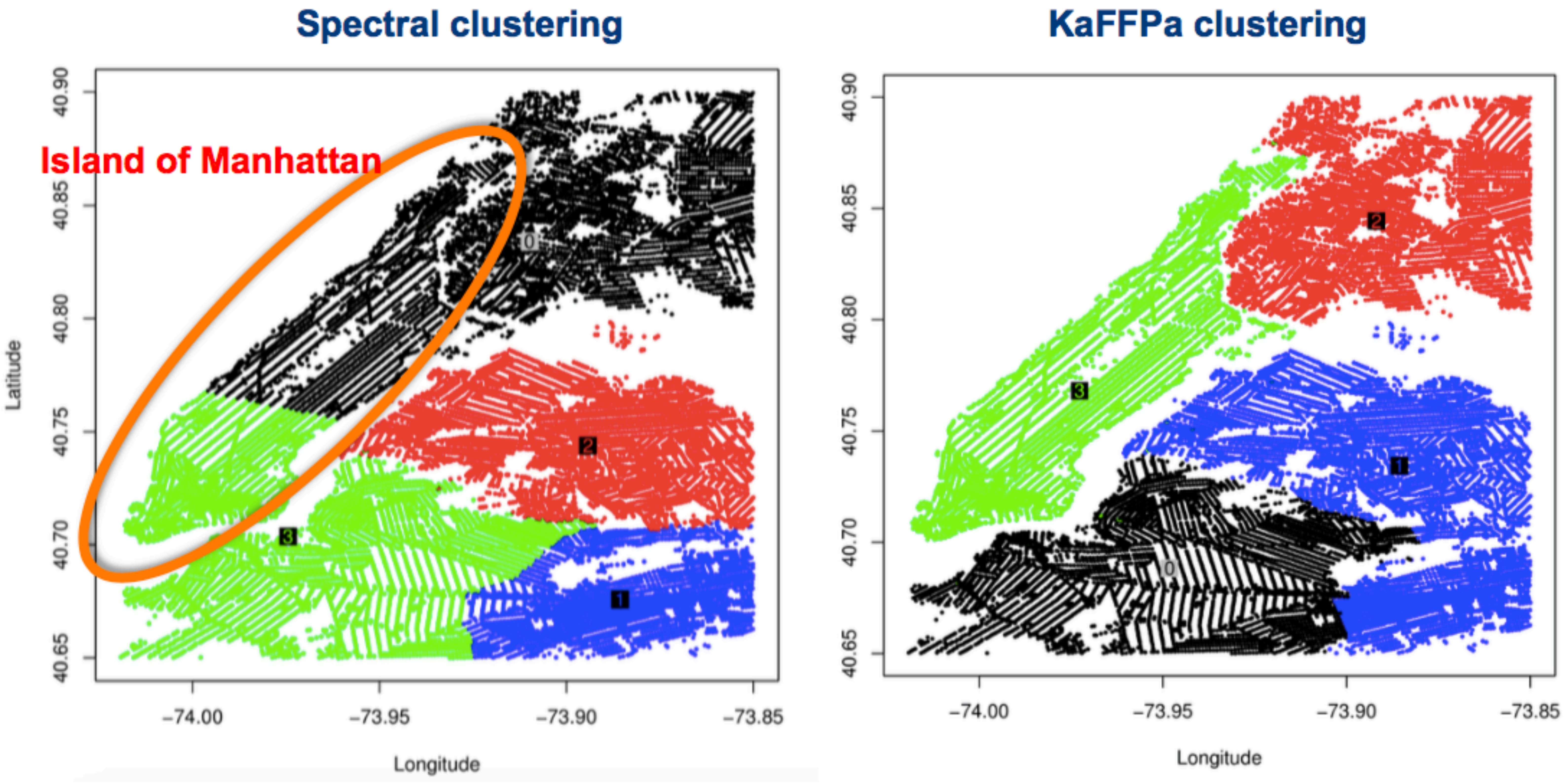


Source: dilbert.com



Example: Comparison between KaFFPa and spectral clustering

illinois.edu



Source: Dan Work (UIUC), Donovan et al. (2016)

27

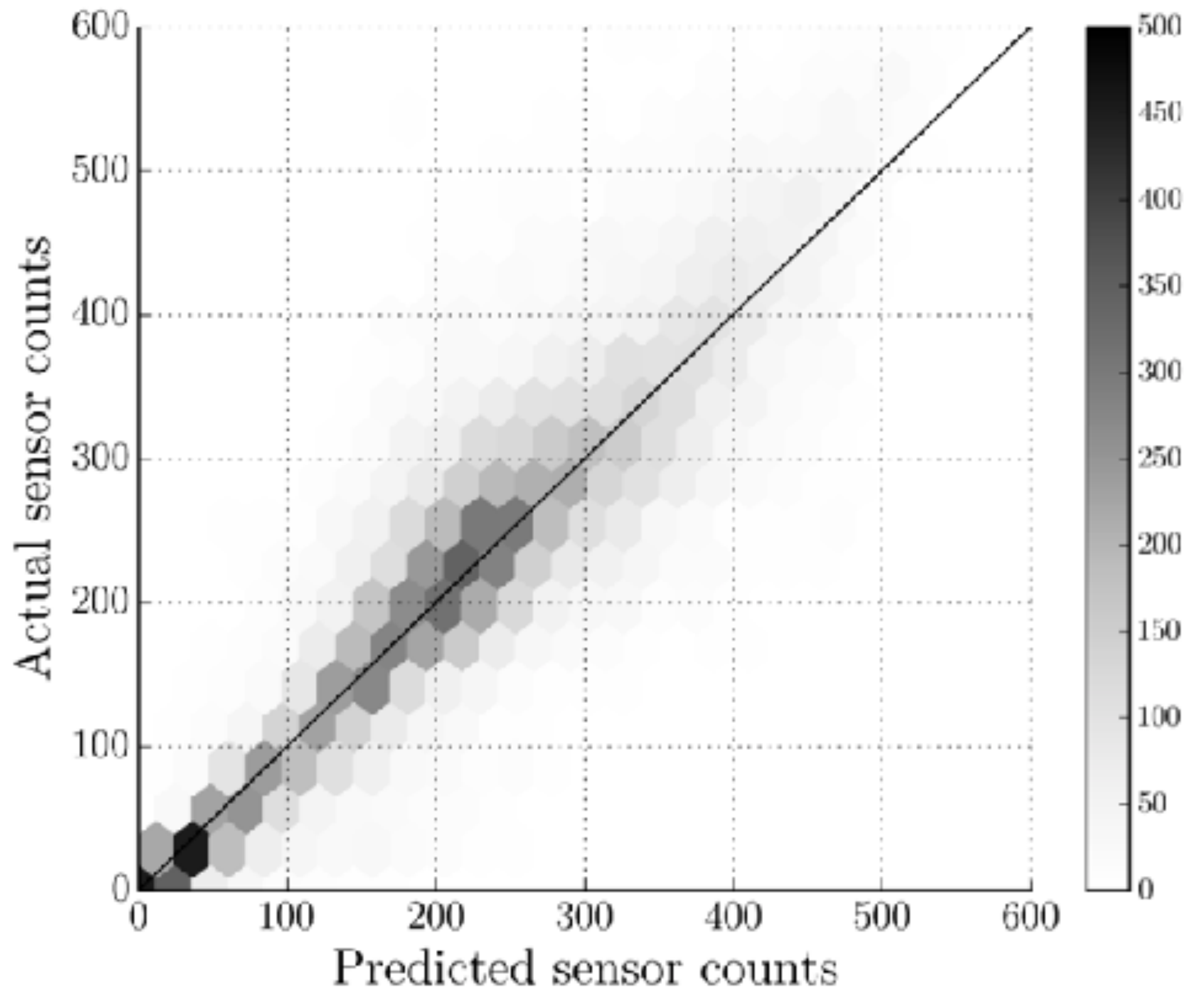
Lehrstuhl für Vernetzte Verkehrssysteme
Fakultät für Bau, Geo und Umwelt
Technische Universität München

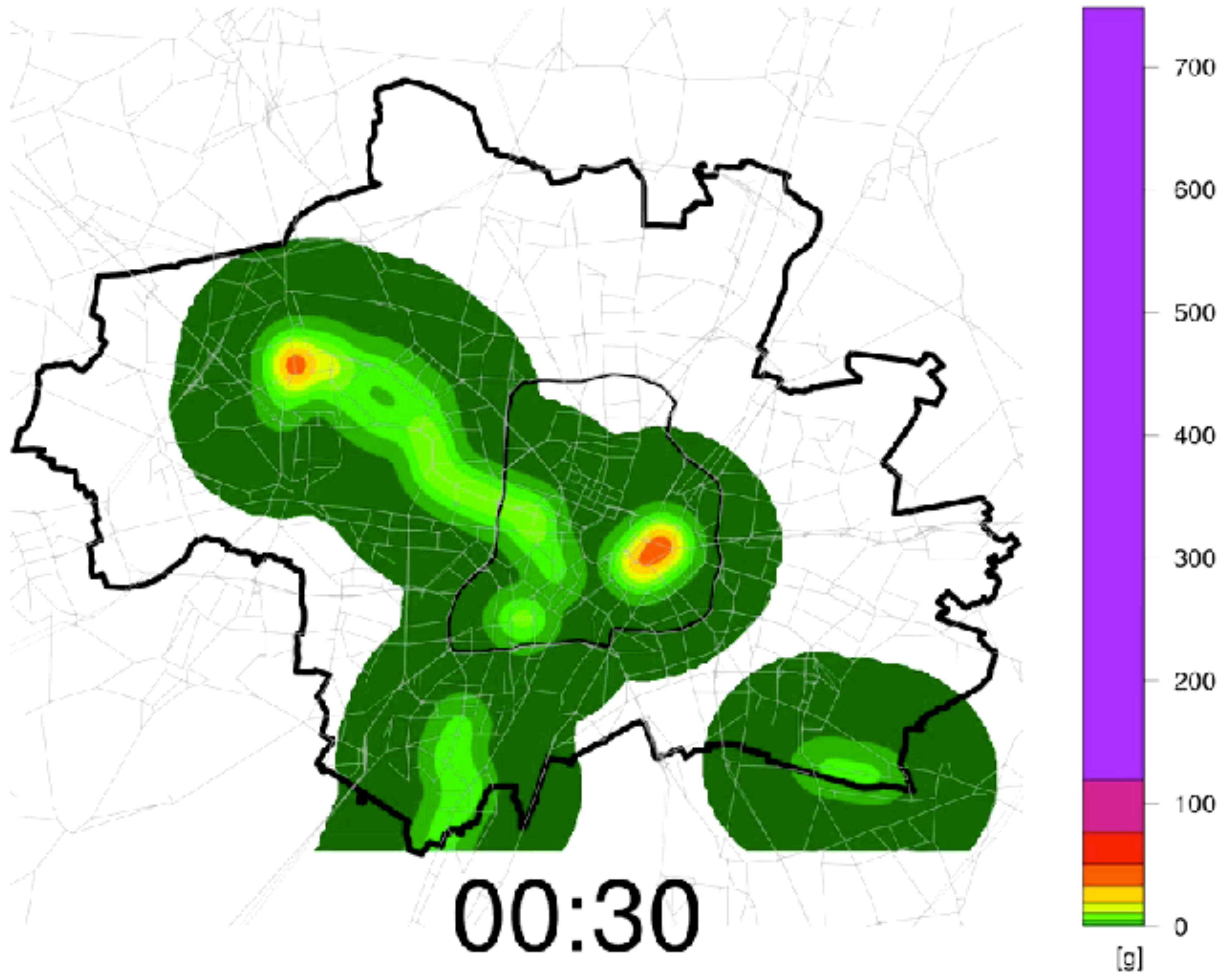
Visualisation

Traditionally it was “easy” to look at the model inputs and outputs

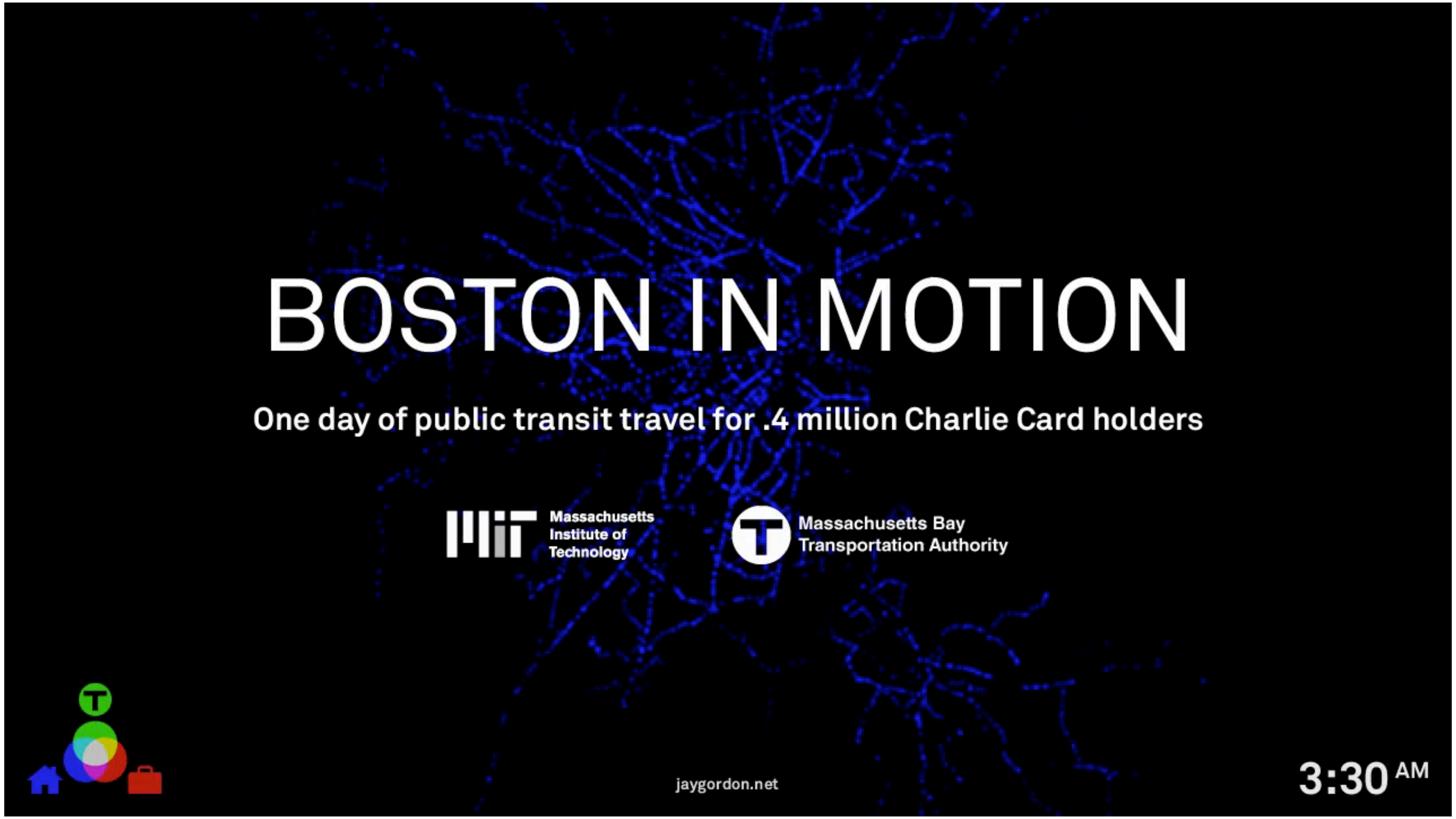
Interpretation and analysis

To understand Big data we need a lot of work and the development of new strategies





Simple visualisations



Lehrstuhl für Vernetzte Verkehrssysteme
Fakultät für Bau, Geo und Umwelt
Technische Universität München

Virtual / Augmented Reality

CAVE (no need for glasses)

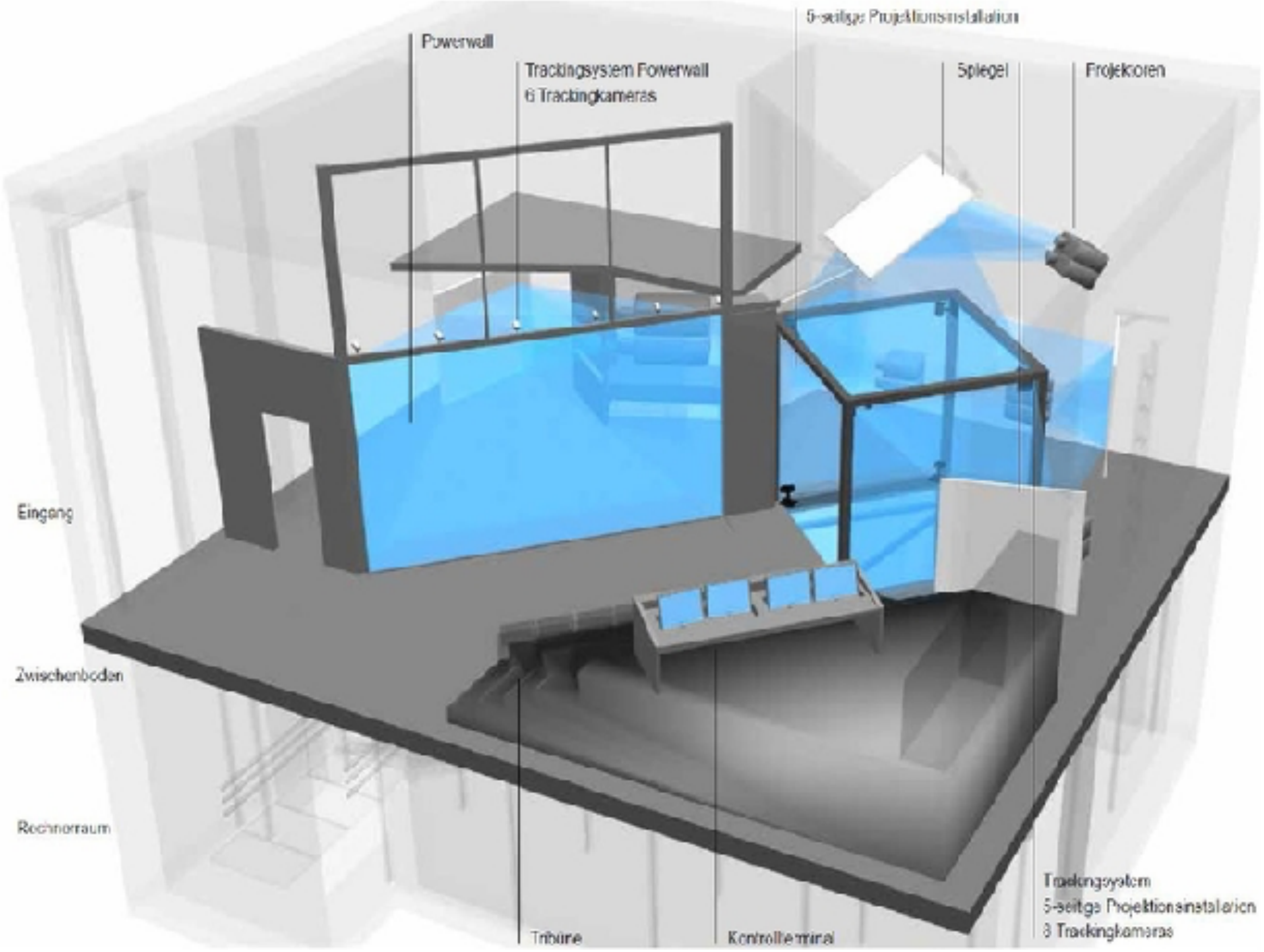
LRZ Virtual Reality and Visualisation Centre (V2C)

LRZ Holobench

More accessible technologies

Oculus Rift, etc.

Upcoming versions will not require powerful computer



Lehrstuhl für Vernetzte Verkehrssysteme
Fakultät für Bau, Geo und Umwelt
Technische Universität München

Privacy: lost opportunities?

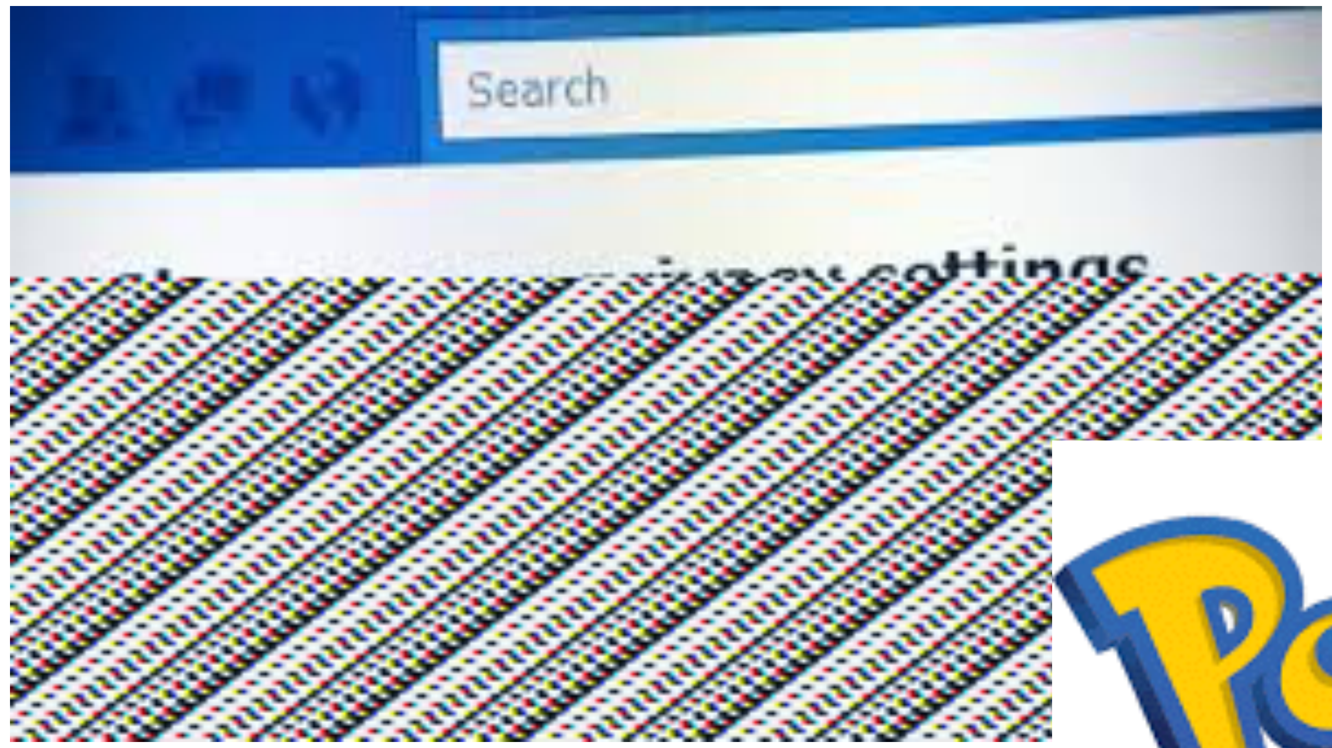
White House Big Data Survey

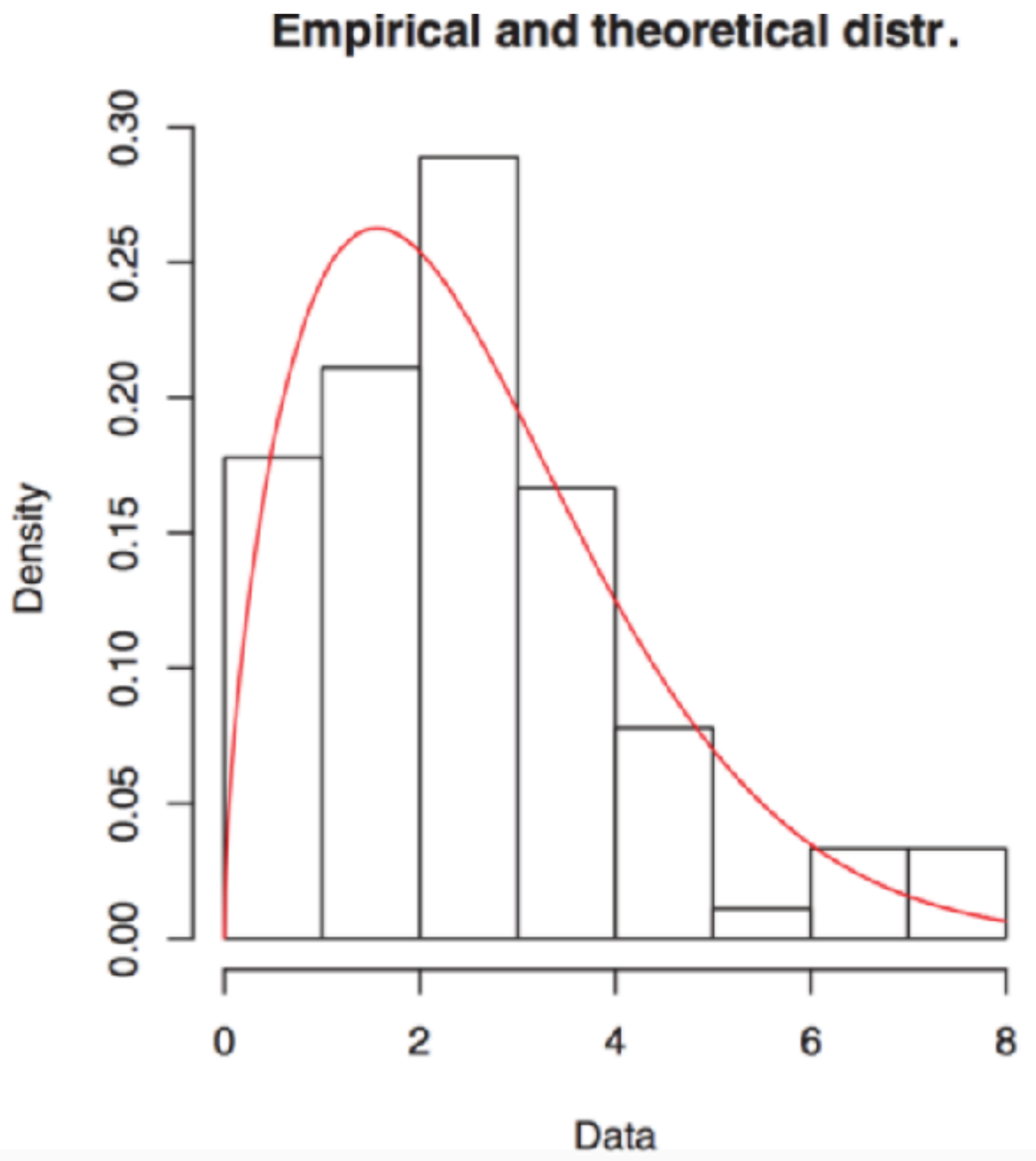
(White House, May 2014)

McKinsey, 2011

Value (and paradox) of privacy

Pokemon Go / Facebook use vs. Privacy concerns

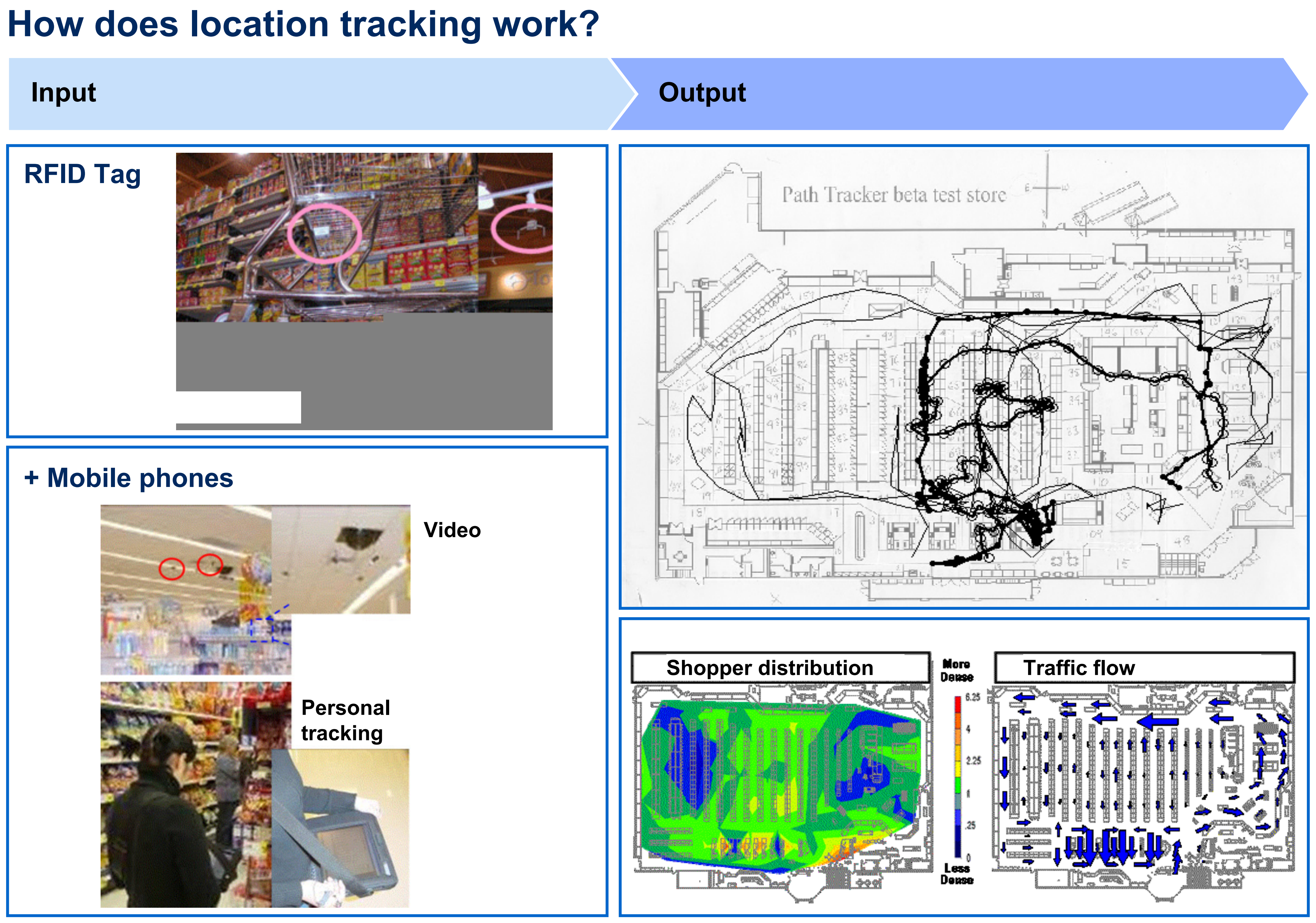




Cost of privacy (€/level)
(Weibull)

Antoniou and Polydoropoulou (2015)

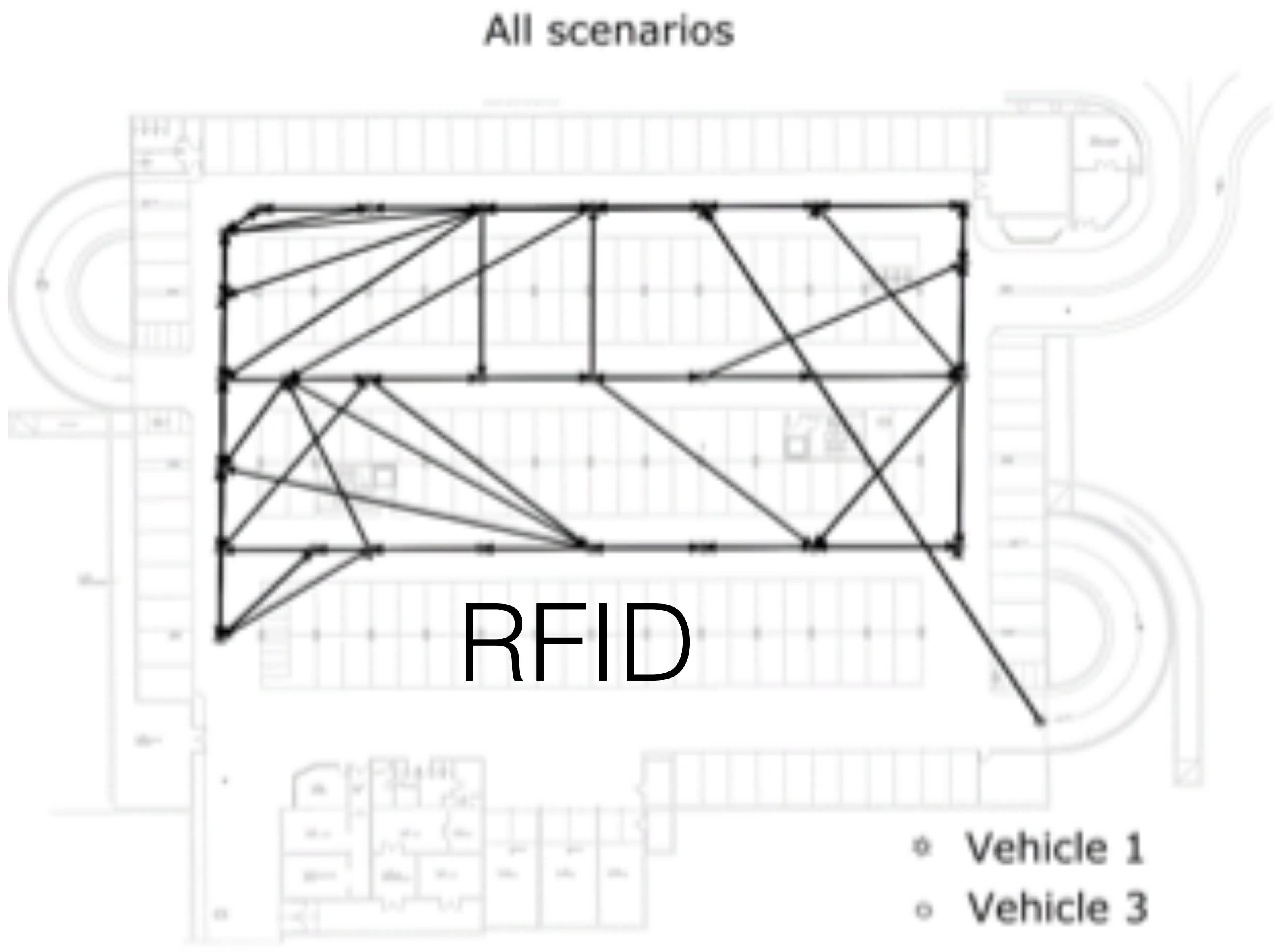
Privacy: lost opportunities?



SOURCE: Press and literature search

McKinsey, 2011

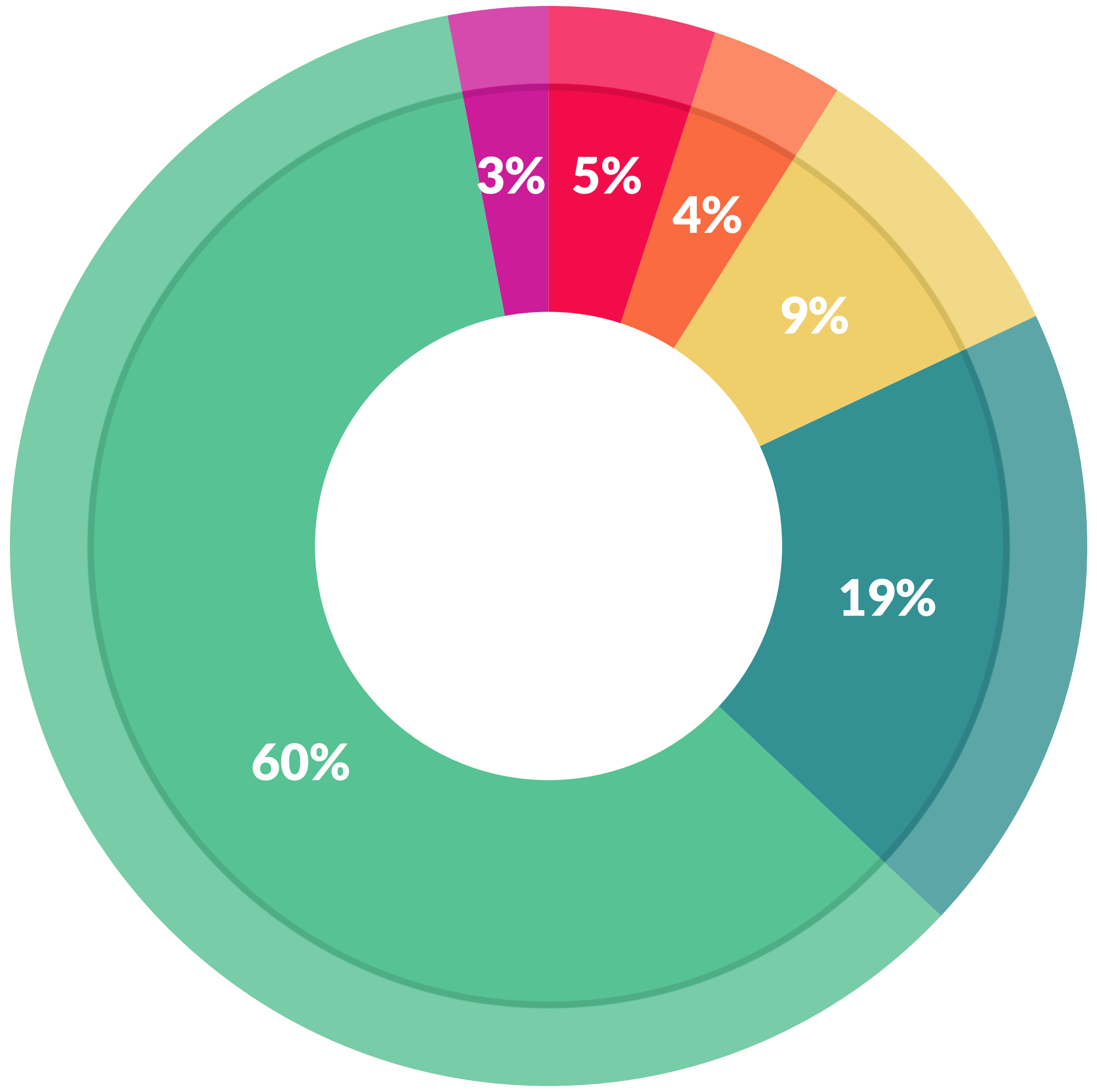
Indoor positioning



UWB

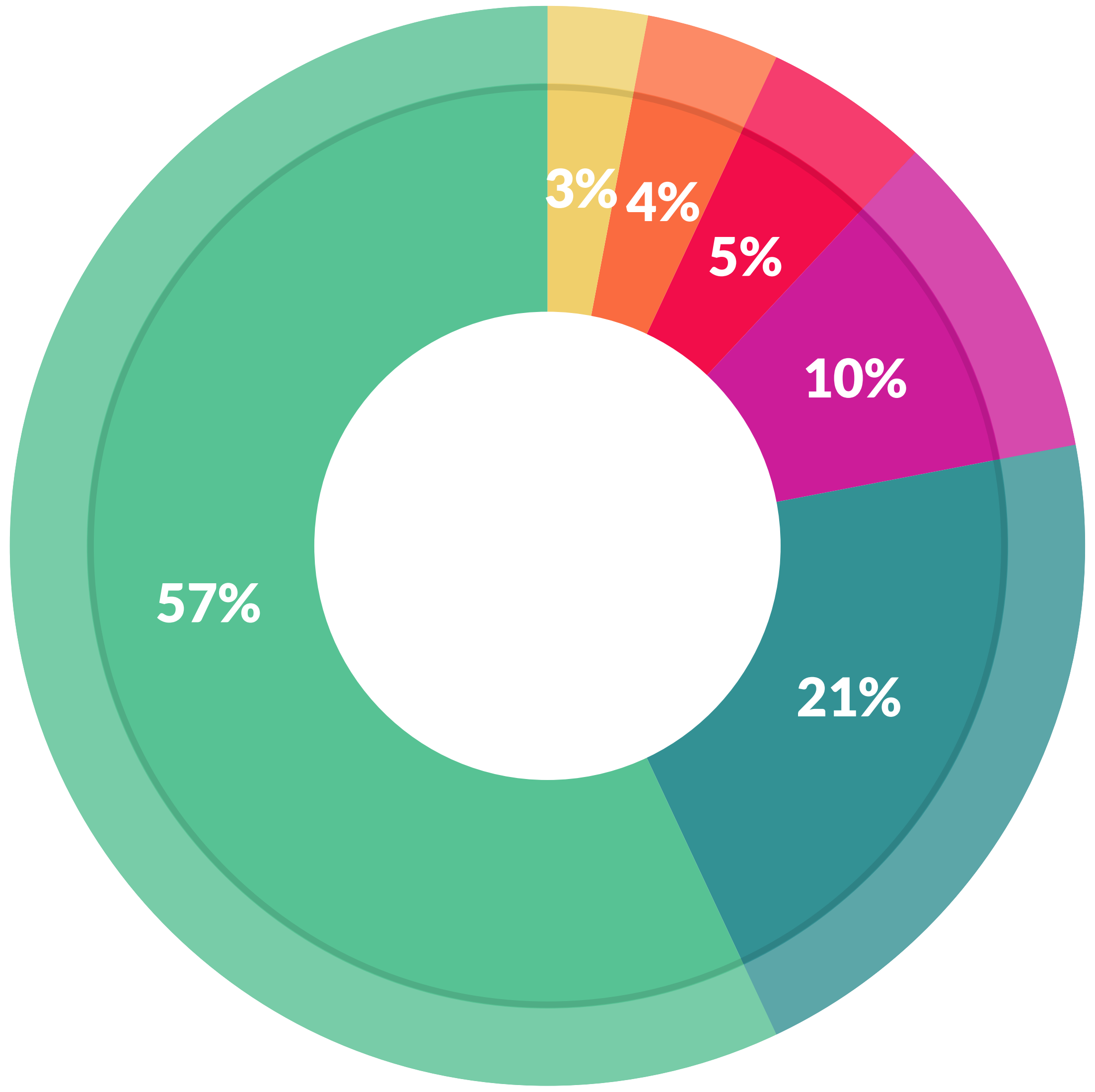
Antoniou et al. 2017

Data analysts - the bad news



What data scientists spend the most time doing

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

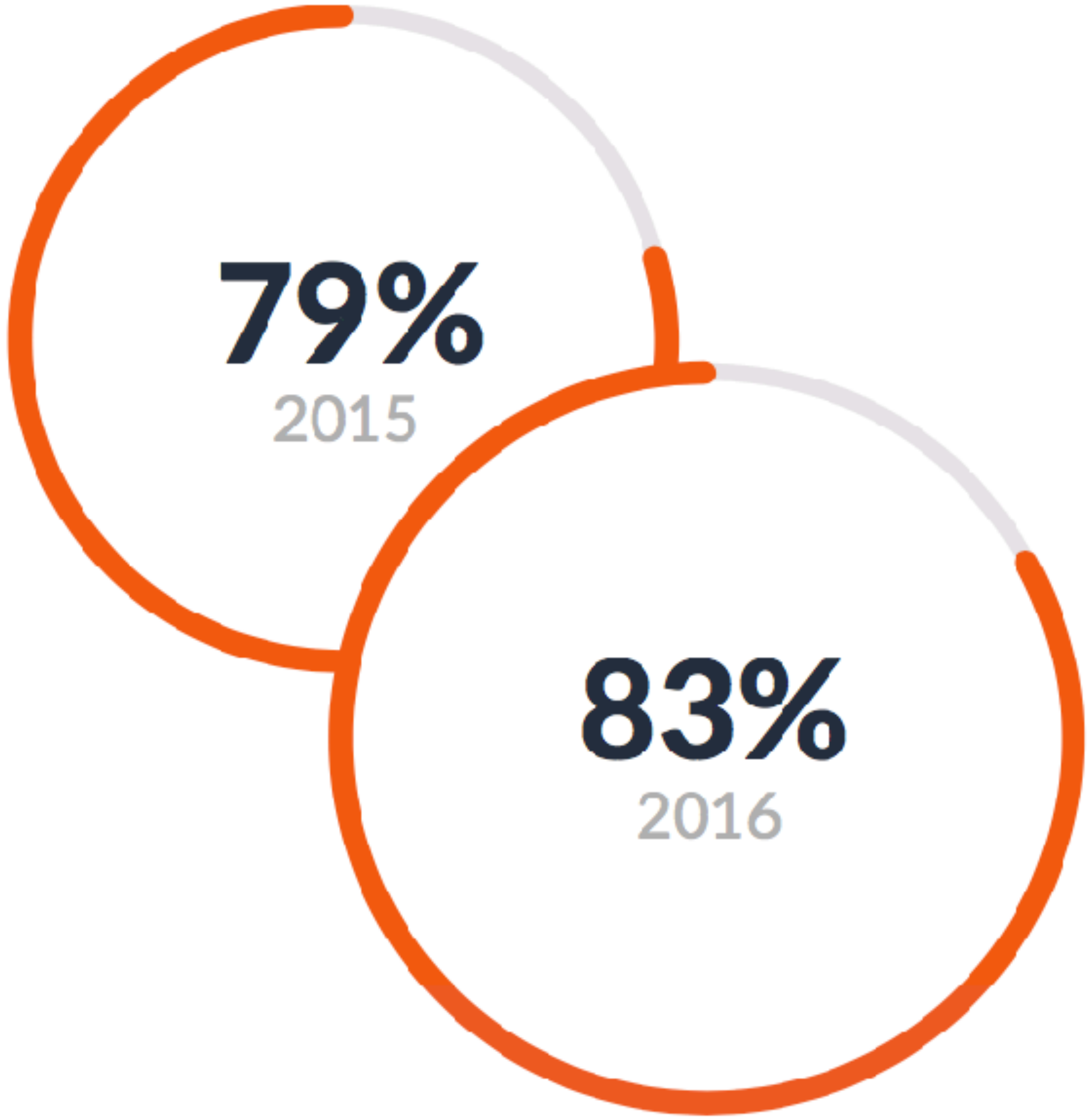


What's the least enjoyable part of data science?

- *Building training sets: 10%*
- *Cleaning and organizing data: 57%*
- *Collecting data sets: 21%*
- *Mining data for patterns: 3%*
- *Refining algorithms: 4%*
- *Other: 5%*

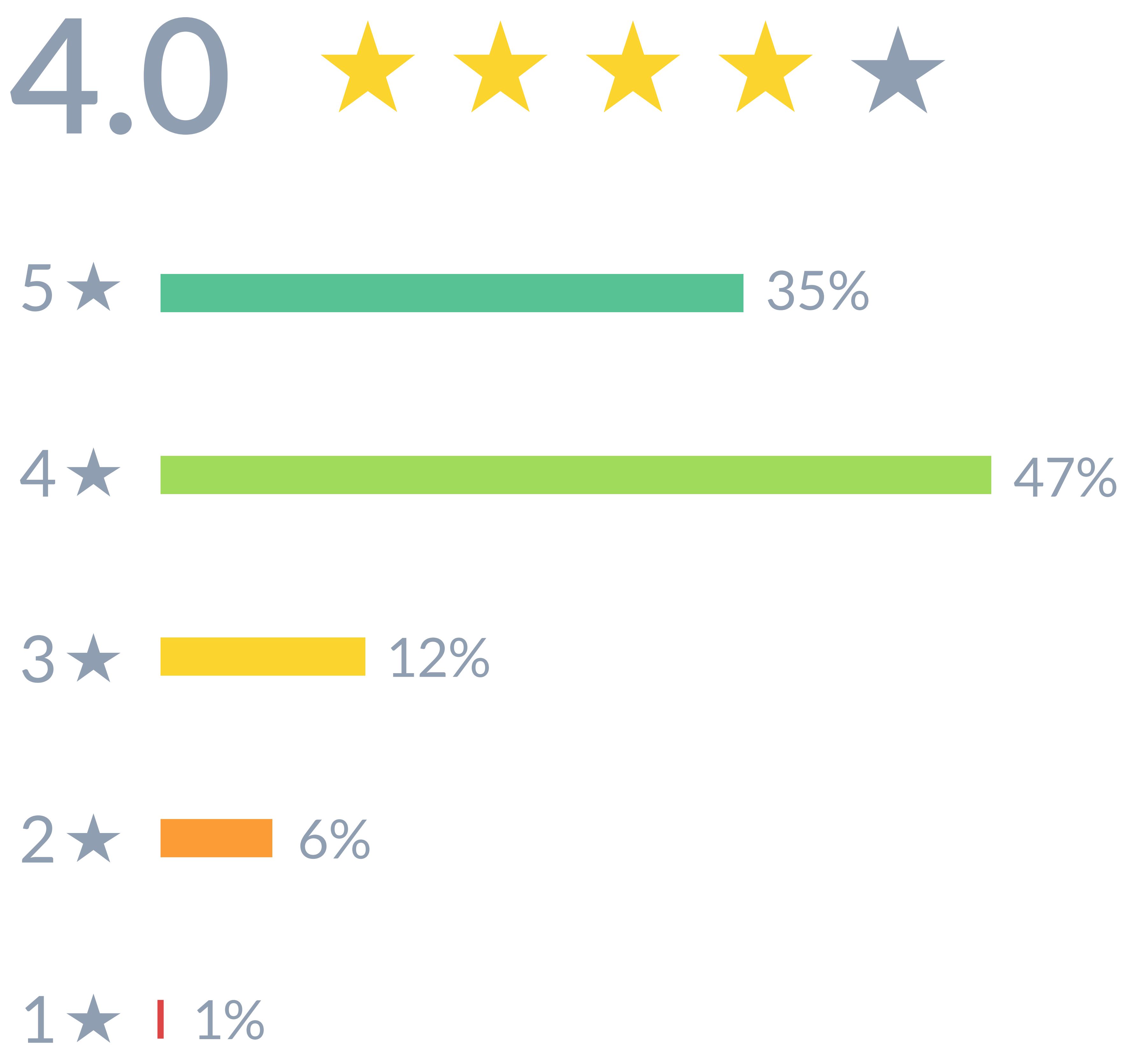
CrowdFlower

Data analysts - the “good” news



Respondents who said there weren’t enough data scientists to go around

CrowdFlower



Data scientist job satisfaction

Need for institutional support

“Big data will take **2 years+ to have effect**.
So, it will be dead in the water,
unless you **get top level management involved**”.

Professor Bjarne Kjaer Ersbøll, DTU Compute
(Copenhagen, 12.10.2016)

References

- Antoniou, C., C. L. Azevedo, L. Lu, F. Pereira and M. Ben-Akiva (2015). W-SPSA in practice: Approximation of weight matrices and calibration of traffic simulation models. Transportation Research Part C: Emerging Technologies, Vol. 59, October, pp. 129-146.
- Antoniou, C. and H. N. Koutsopoulos. Estimation of Traffic Dynamics Models with Machine Learning Methods. Transportation Research Record: Journal of the Transportation Research Board, No. 1965, pp. 103-111, Washington D.C., 2006.
- Antoniou, C., H.N. Koutsopoulos and G.Yannis (2013). Dynamic Data-Driven Local Traffic State Estimation and Prediction. Transportation Research Part C: Emerging Technologies, Vol. 34, pp. 89-107.
- Antoniou, C. and A. Polydoropoulou (2015). The value of privacy. Evidence from the use of mobile devices for traveler information systems. Journal of Intelligent Transportation Systems: Technology, Planning and Operations, 19(2), pp. 167-180.
- Arun Prakash, A., R. Seshardi, C. Antoniou, F. Pereira and M. Ben-Akiva (2017). Reducing the dimension of online calibration in Dynamic Traffic Assignment systems, 96th Annual Meeting of the Transportation Research Board, January 8-12, 2017, Washington, D.C.
- Chaniotakis, E., C. Antoniou, G. Aifadopoulou, and L. Dimitriou (2017). Inferring activities from social media data. 96th Annual Meeting of the Transportation Research Board, January 8-12, 2017, Washington, D.C.
- Chaniotakis E., C. Antoniou, J. M. Salanova Grau, L. Dimitriou (2016). Can Social Media Data Augment Travel Demand Survey Data? 19th IEEE Intelligent Transportation Systems Conference, November 1-4, 2016, Rio de Janeiro, Brazil.
- Chaniotakis, E., C. Antoniou and F. Pereira (2016). Mapping Social Media for Transportation Studies, IEEE Intelligent Systems Magazine.
- CrowdFlower (2016). 2016 Data Science Report, available online at crowdflower.com
- Donovan, B., J. Lee, and D. Work (2016). “A high resolution method for quantifying resilience of urban road networks.” submitted to IEEE Transactions on Intelligent Transportation Systems.
- Papathanasopoulou, V., I. Markou and C. Antoniou (2016). Online calibration for microscopic traffic simulation and dynamic multi-step prediction of traffic speed. Transportation Research Part C: Emerging Technologies, 68, pp. 144-159.

Big data for transportation systems analysis
Challenges and opportunities

Univ.-Prof. Dr. Constantinos Antoniou

Technische Universität München

Fakultät für Bau, Geo und Umwelt

Lehrstuhl für Vernetzte Verkehrssysteme

c.antoniou@tum.de

Graz, 17. Mai 2018