Non-recurrent events in traffic prediction

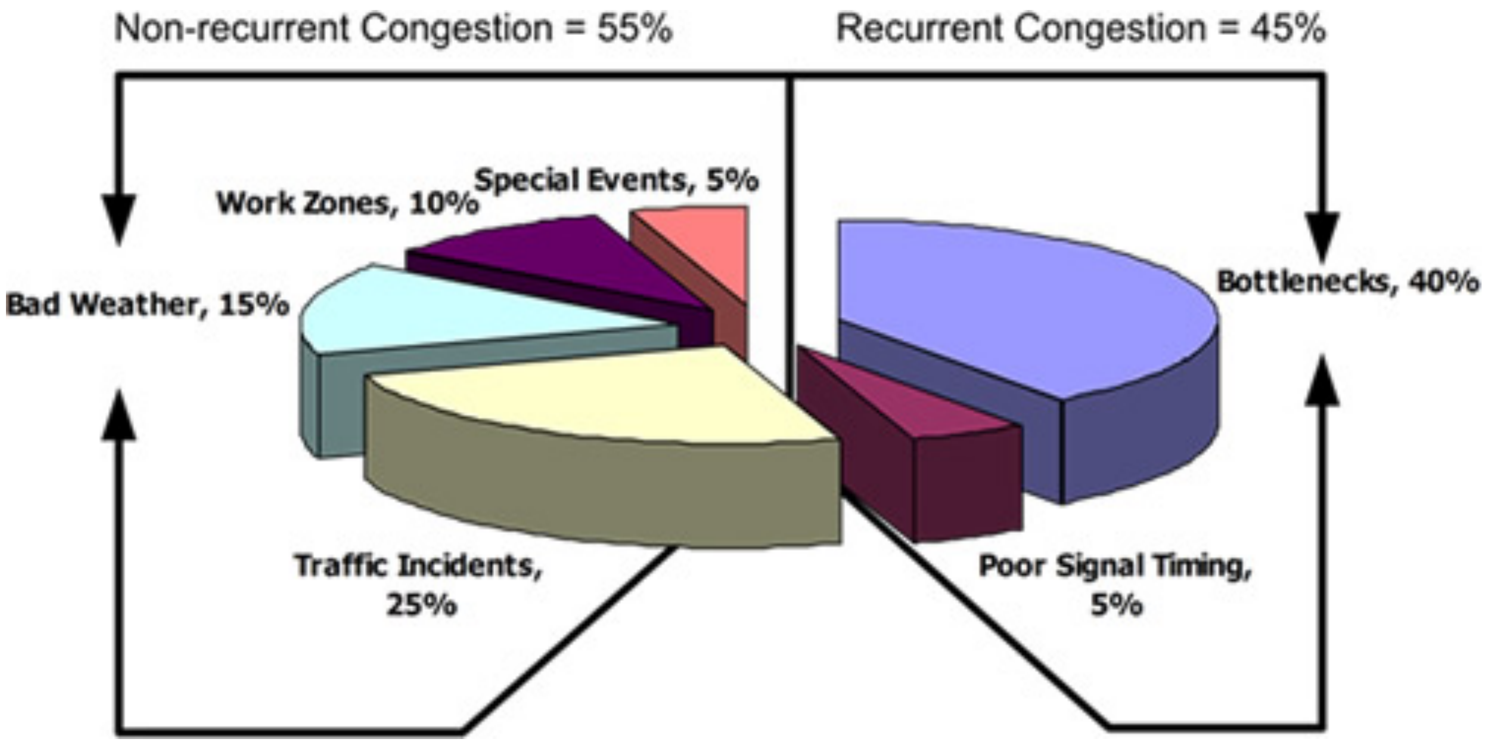
Francisco Pereira

IPAM Workshop IV

17th November 2015

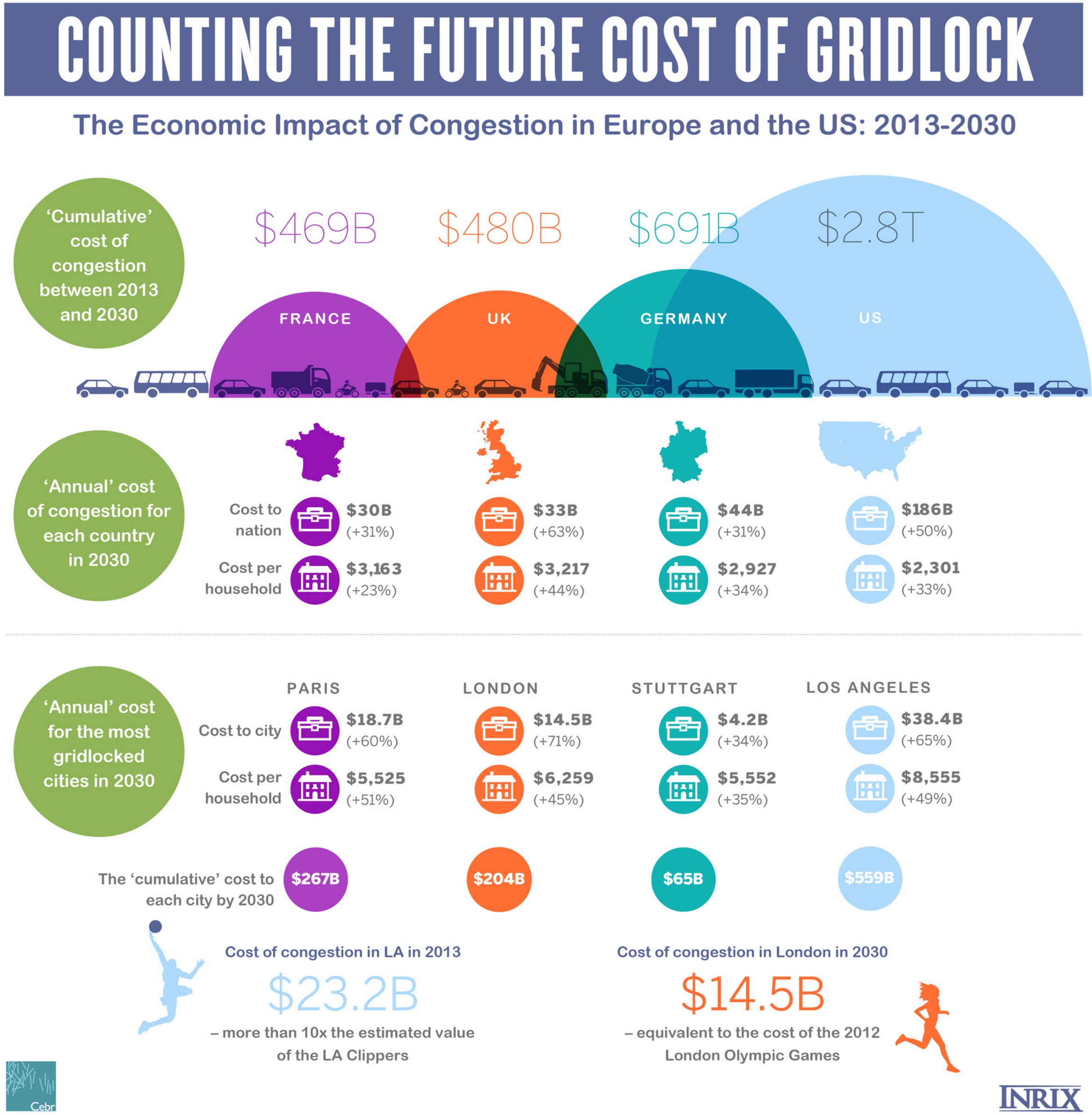
Collaborations: MIT, Singapore (SMART), Filipe Rodrigues, Stanislav Borysov, Aidan O'Sullivan, Haizheng Zhang, Constantinos Antoniou, Moshe Ben-Akiva

The cost of disruptions



Source: FHWA, 2005

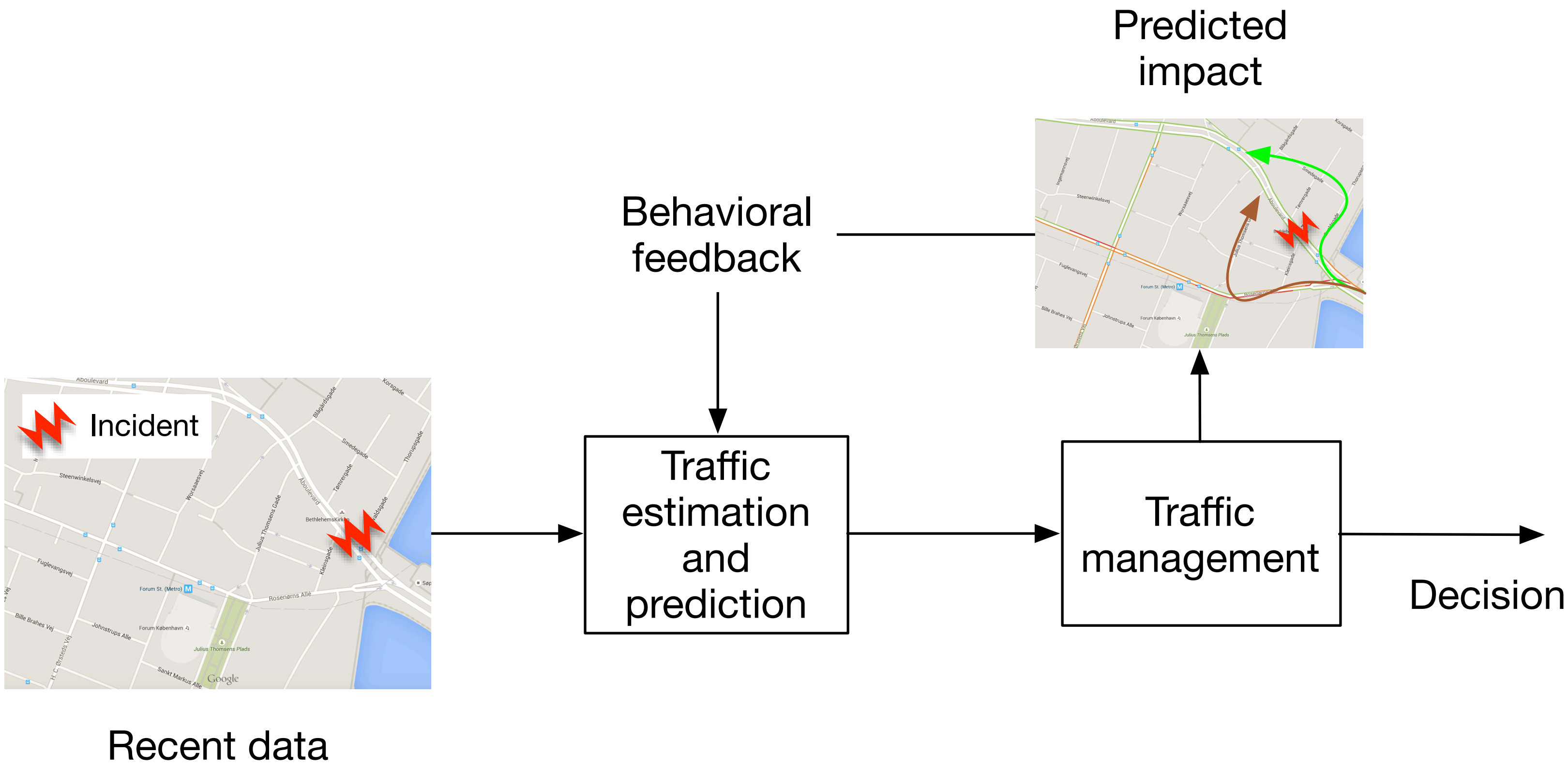
The cost of disruptions



$187B/year in USA

$102.8B/yr in non-recurrent congestion!

Predictive traffic management



…but do we really have a problem?

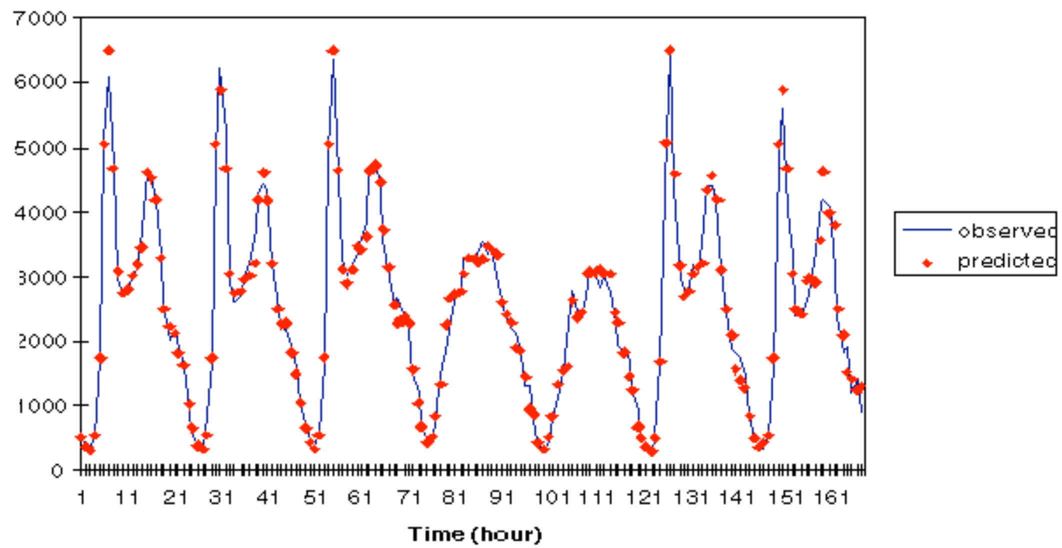


Figure 7a- Traffic flow (actual compared to the predicted value) at sensor station 605 EB for the first week of March 1995

From http://rodin.wustl.edu/cosc/doc1.html

Traffic prediction

In fact, the problem is **generally solved** (for a long time) for recurrent or habitual circumstances

But…

. Under recurrent/habitual behavior, we already know what to expect!

. Help is welcome under unexpected circumstances…

Two approaches to traffic prediction

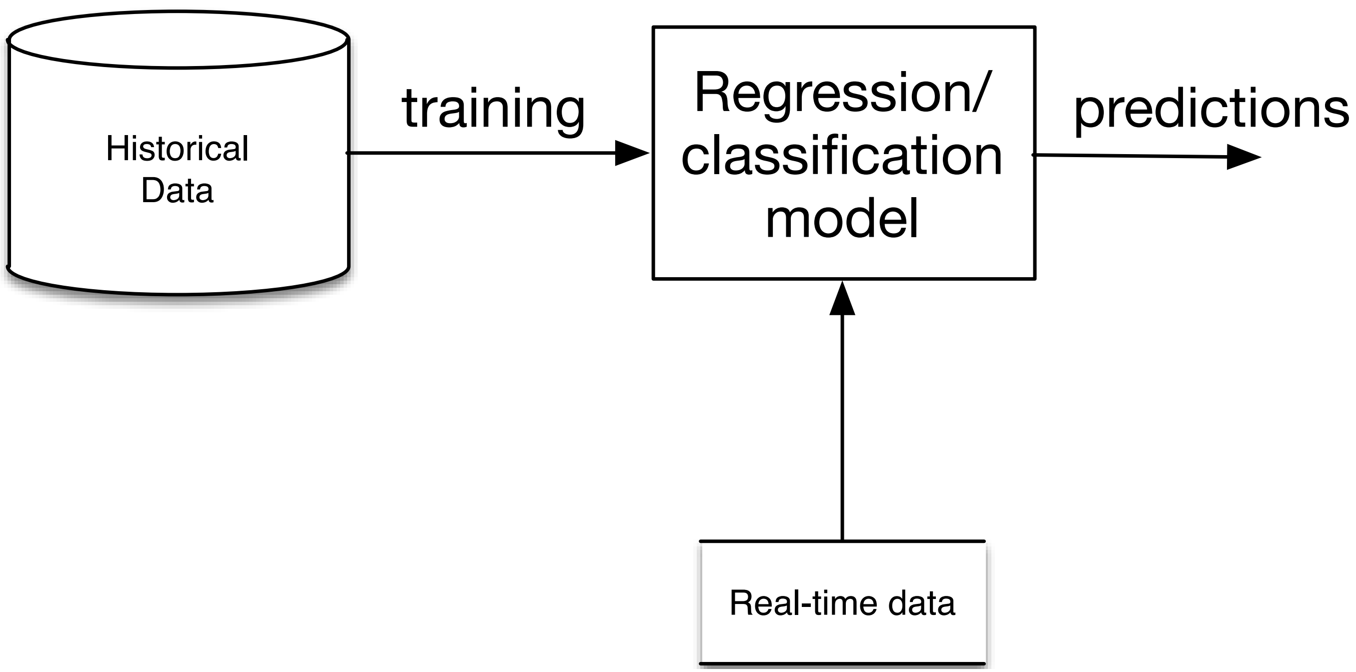
. Data driven

. Model based

Two approaches to traffic prediction

Data driven

- Mostly statistical (time series, neural networks,…)
- Based on historical data – trouble with ***non-recurrent events***!
- Computationally efficient



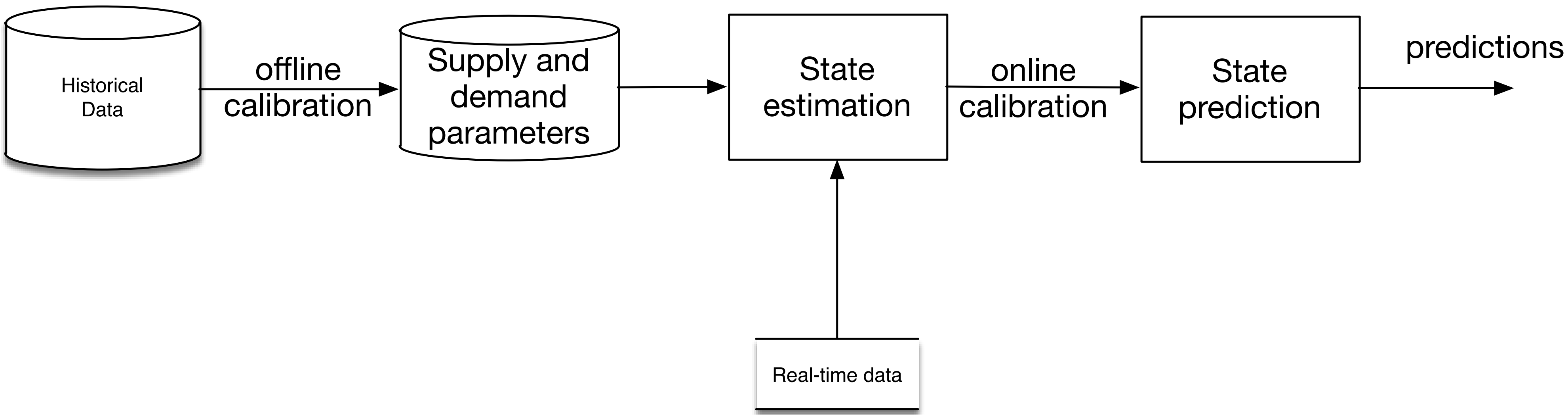
Two approaches to traffic prediction

Model based

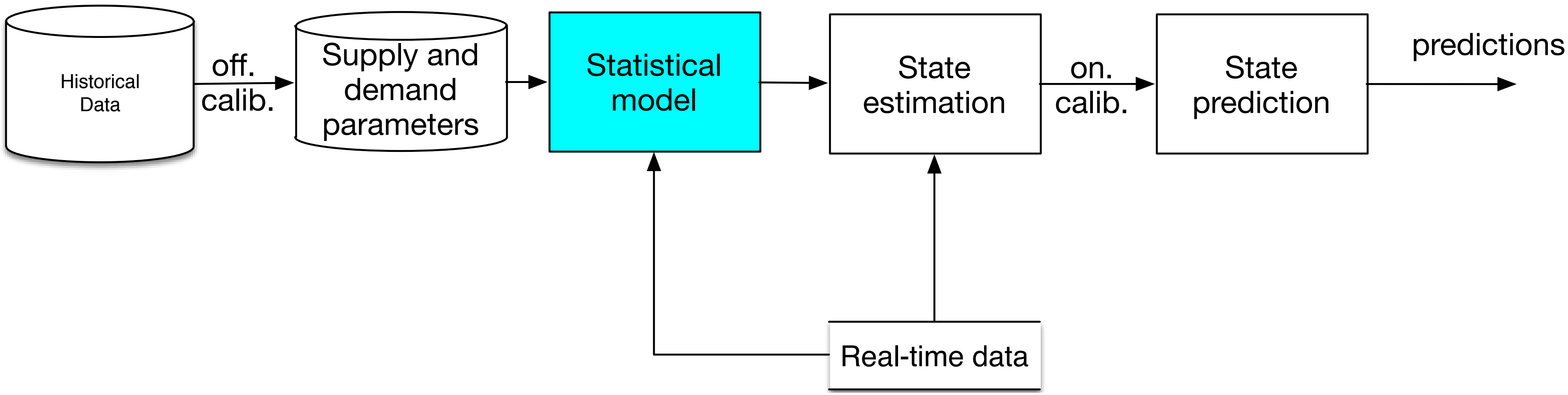
. Mostly simulation based (mesoscopic level)

. Represent individual driving behavior (path choice, departure time choice)

. Computationally heavy – trouble with ***non-recurrent events***!



Best of both worlds?



Which parameters?

Demand

. Origin/destination matrix

. Behavior models

Supply

. Capacity reduction

. Speed density functions

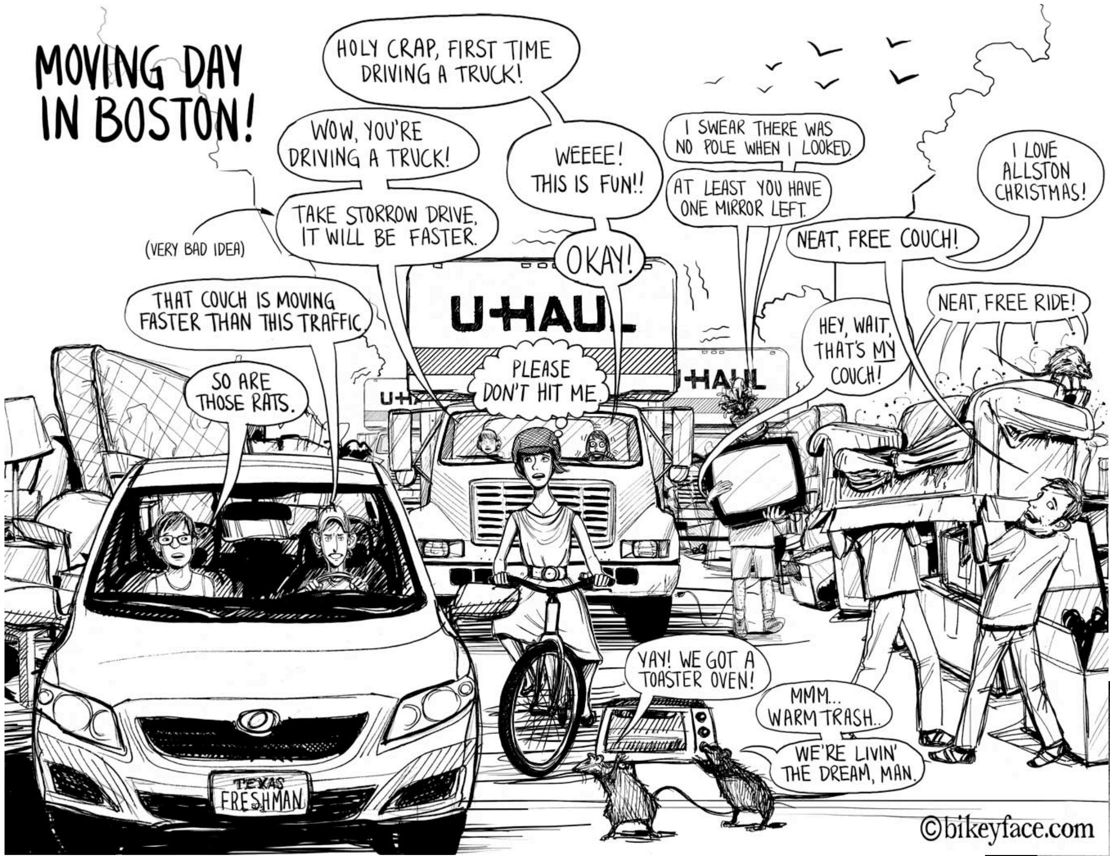
Non-recurrent events

Demand	Supply	
		Predictable
		Non predictable

Non-recurrent events

Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	**Predictable**
crisis situations, unknown high fluctuations	incidents, weather, crisis situations	**Non predictable**

Predictable demand



Predictable demand

Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	**Predictable**
crisis situations, unknown high fluctuations	incidents, weather, crisis situations	**Non predictable**

Predictable demand

Relevant events on the internet

. Announced by organizers (e.g. events websites)

. Communicated in social networks (e.g. twitter, facebook)

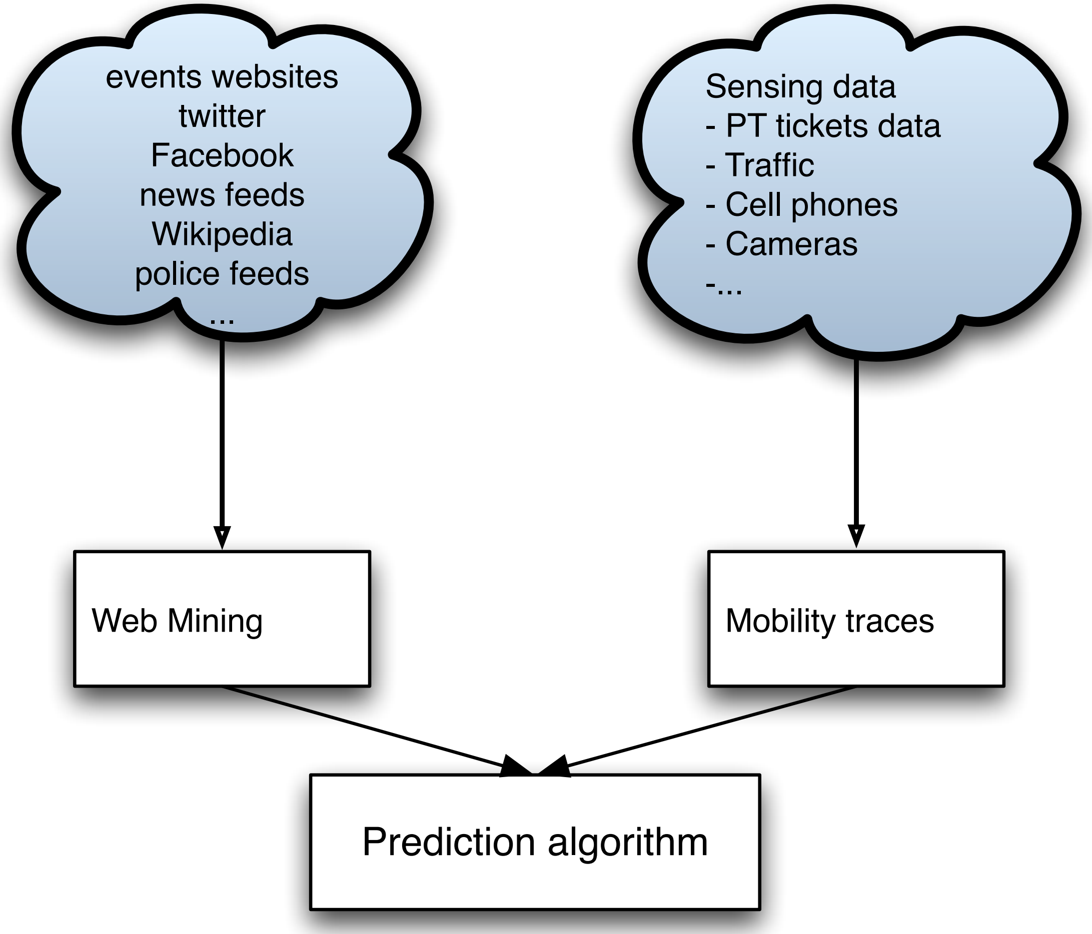
. Schools and official resources

. Religious venues websites/social networks

Lots of potential sources, but **a lot in free form text**

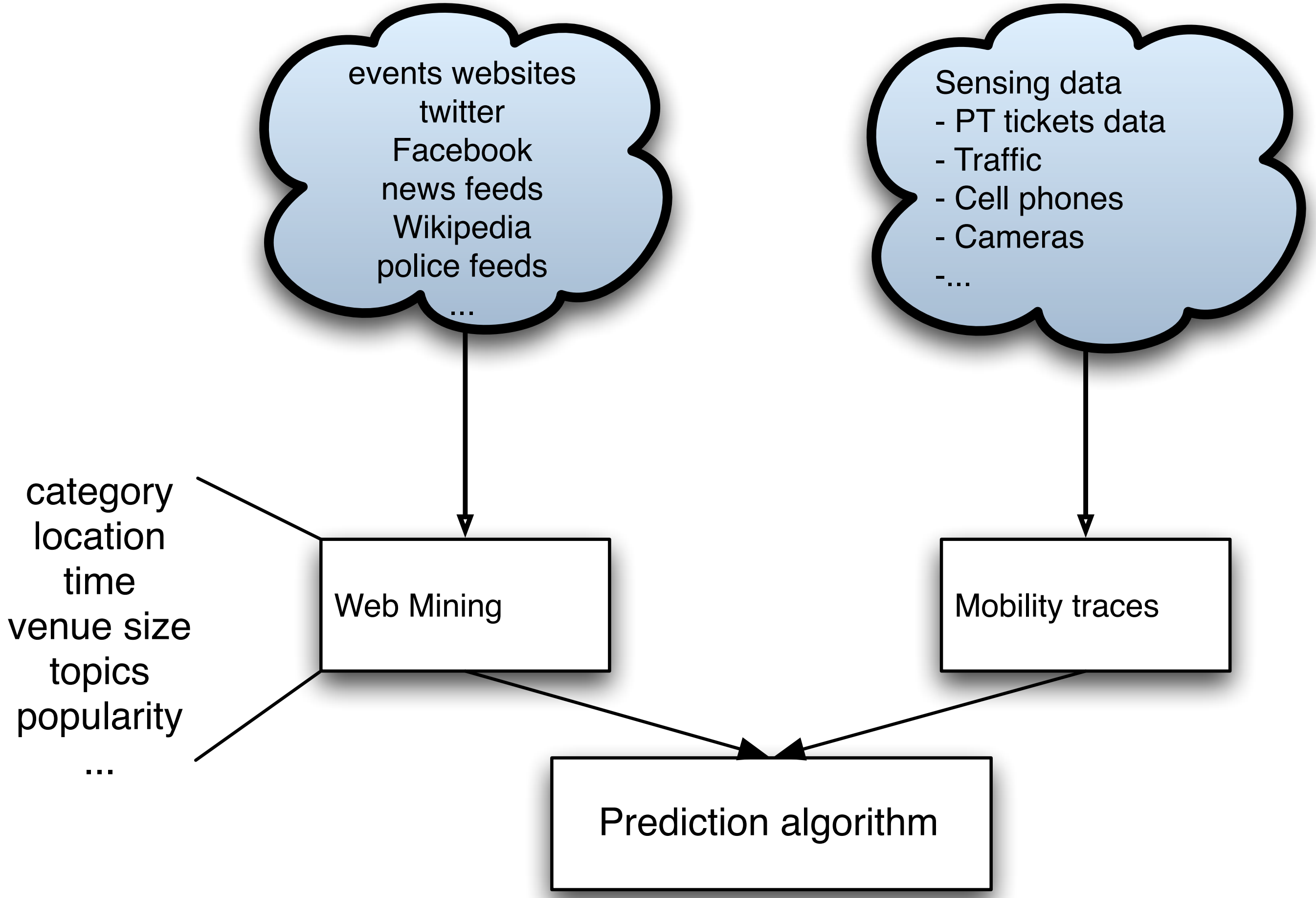
Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	**Predictable**
crisis situations, unknown high fluctuations	Incidents, weather, crisis situations	**Nonpredictable**

Predictable demand



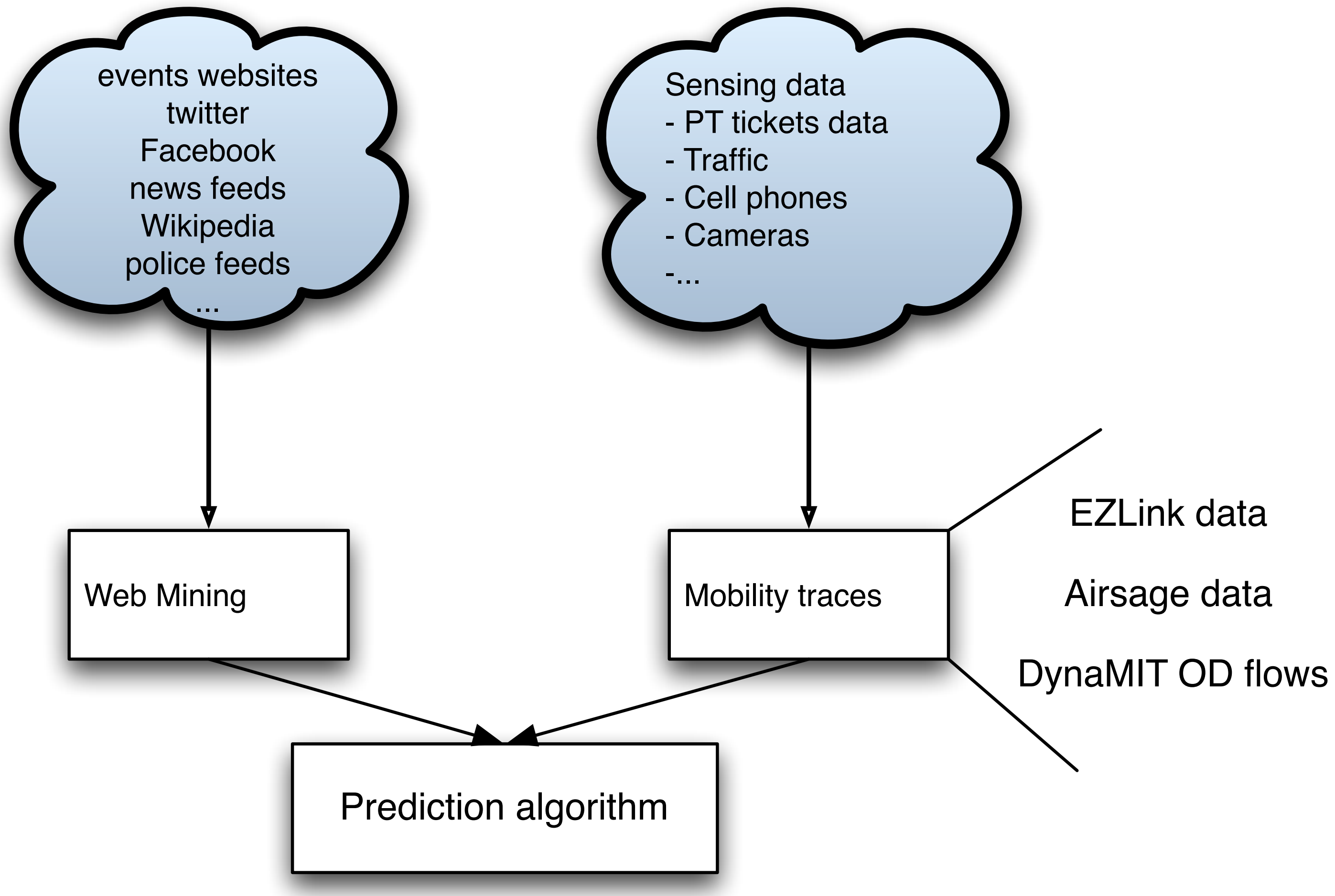
Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	**Predictable**
crisis situations, unknown high fluctuations	Incidents, weather, crisis situations	**Nonpredictable**

Predictable demand



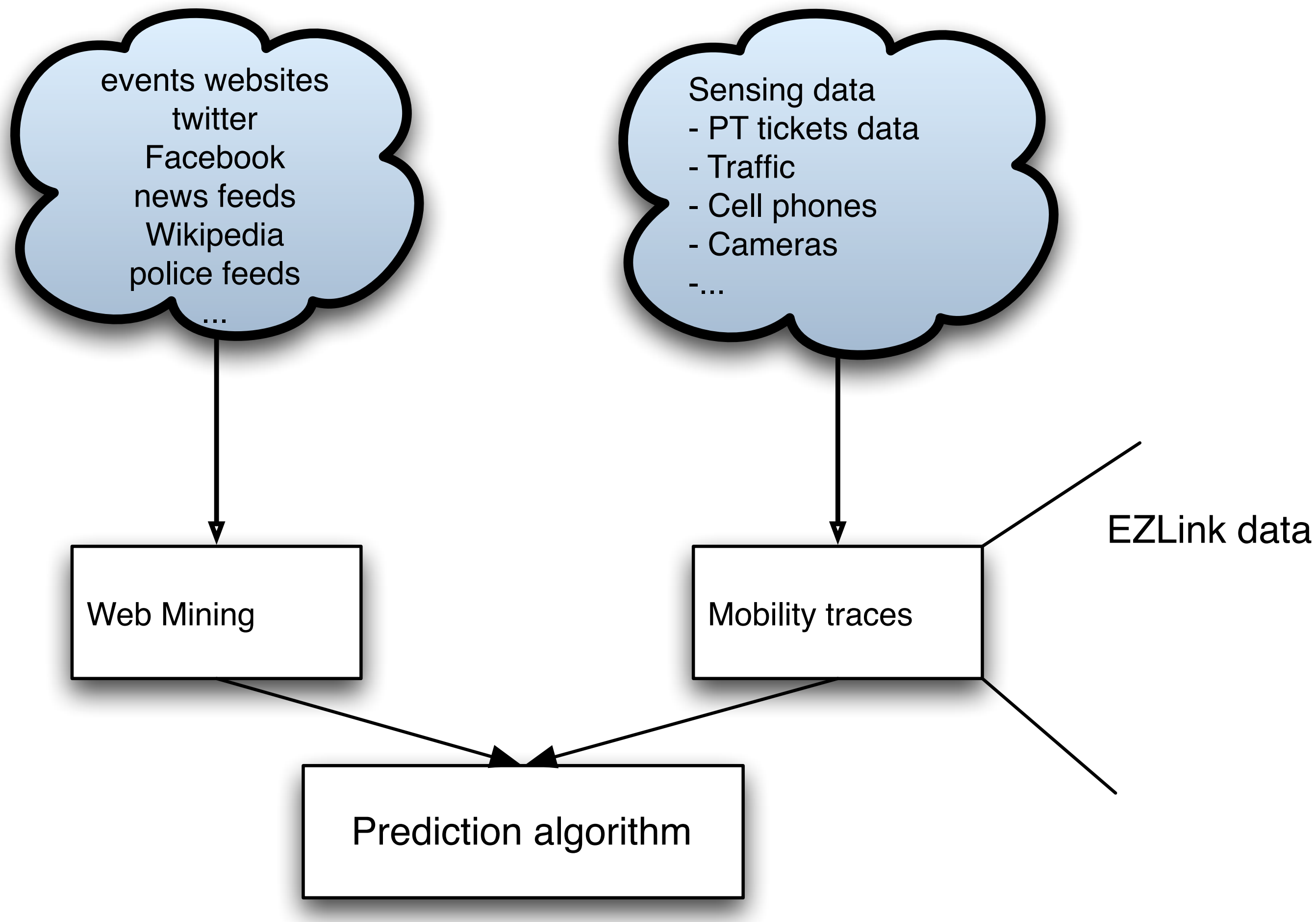
Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	Predictable
crisis situations, unknown high fluctuations	Incidents, weather, crisis situations	Nonpredictable

Predictable demand



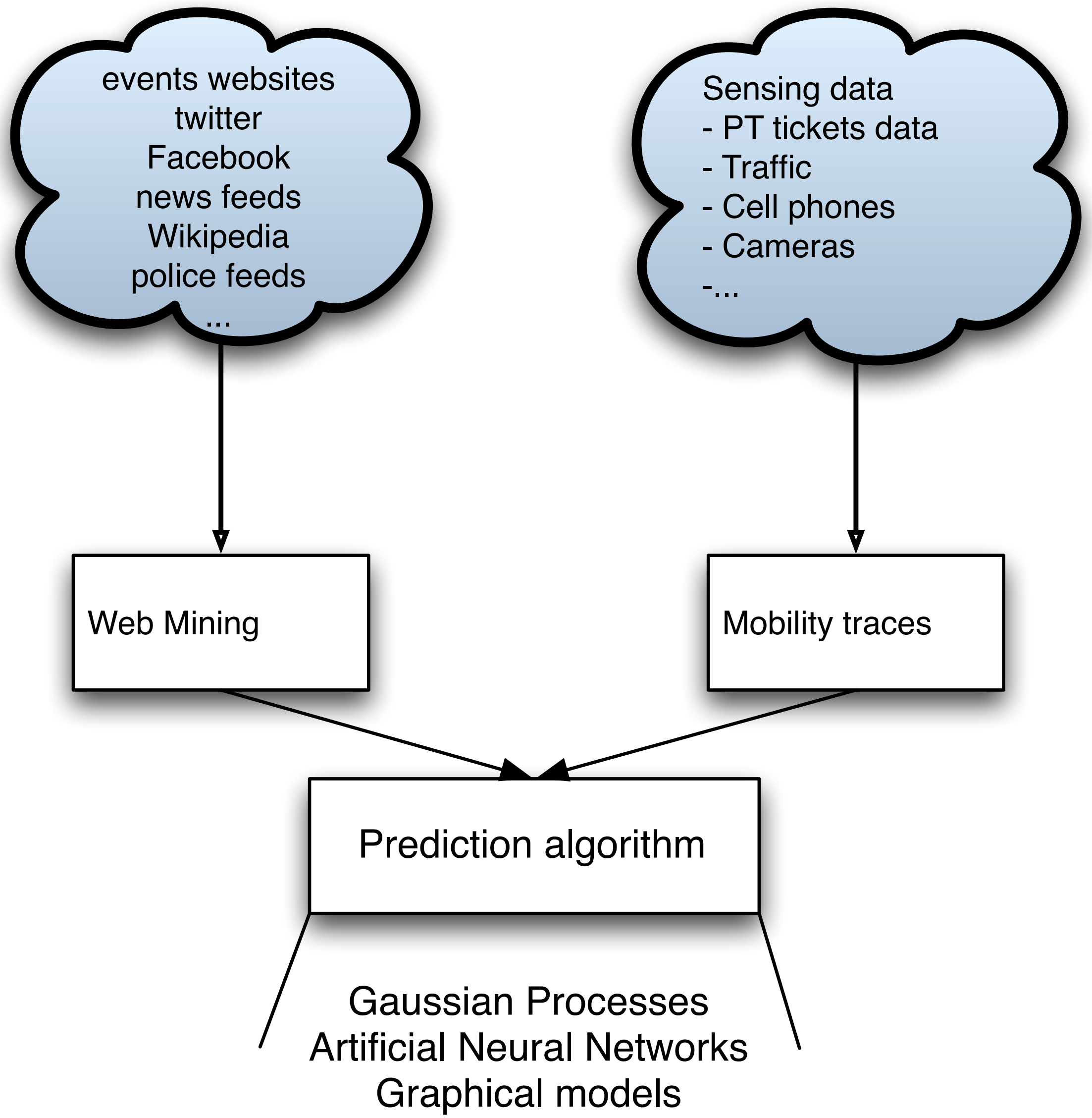
Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	Predictable
crisis situations, unknown high fluctuations	Incidents, weather, crisis situations	Nonpredictable

Predictable demand



Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	**Predictable**
crisis situations, unknown high fluctuations	Incidents, weather, crisis situations	**Nonpredictable**

Predictable demand



Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	**Predictable**
crisis situations, unknown high fluctuations	Incidents, weather, crisis situations	**Nonpredictable**

Modeling demand in planned events

Challenges:

. Multiple **simultaneous** events

. Trip purpose is **not** observed

Assumption:

Total demand, h_n, **is sum of components**

$$h_n = a_n + b_n + \epsilon$$

n – *time interval* ($n<N$)

a_n – *routine component*

b_n – *events component*

ϵ – *error term,*

Modeling demand in planned events

$$h_n = a_n + b_n + \epsilon$$

$$a_n \sim \mathcal{N}(a_n | \boldsymbol{\eta}_a^T \mathbf{x}_n^a, \beta_a)$$

$$b_n = \sum_{i=1}^{E_n} e_n^i, \text{with } e_n^i \sim \mathcal{N}(e_n^i | \boldsymbol{\eta}_e^T \mathbf{x}_n^{e_i}, \beta_e)$$

$$\epsilon \sim \mathcal{N}(\epsilon | 0, v)$$

with e_n^i - demand for event i $(i < E_n)$

$\mathbf{x}_n^a$ - (vector of) routine variables (time of day, day of week…)

$\mathbf{x}_n^{e_i}$ - event variables (discussed shortly…)

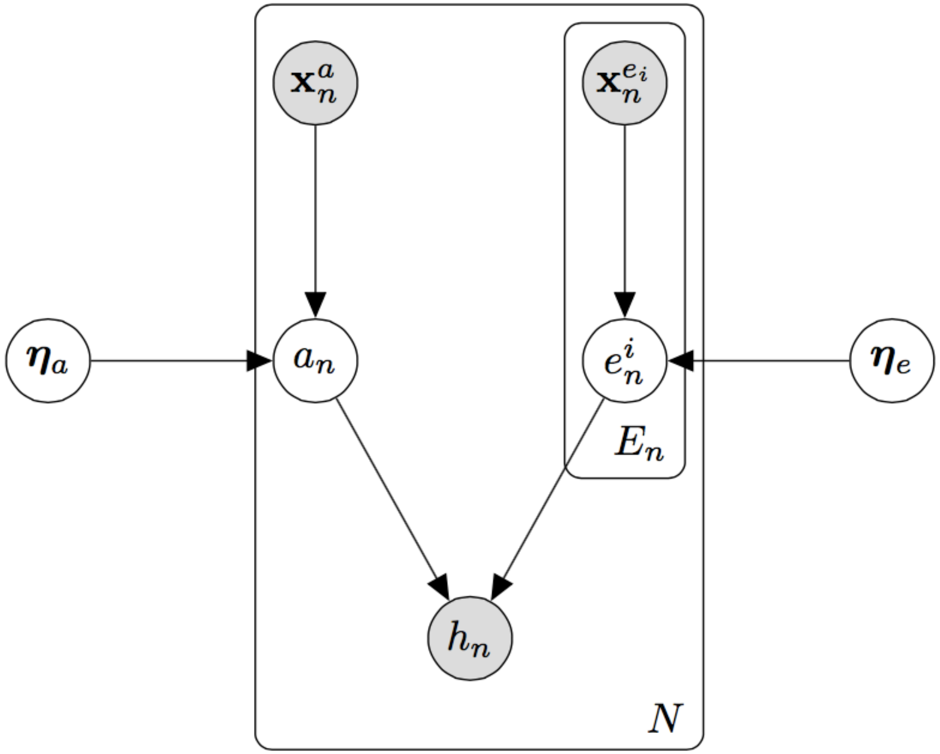
$\boldsymbol{\eta}_a, \boldsymbol{\eta}_e$ - model parameters

v, β_e, β_a – variances (externally defined)

Linear model assumption is illustrative

Modeling demand in planned events

As a generative (graphical) model



Our goal:

$$p(h_n, a_n, \mathbf{e}_n | \boldsymbol{\eta}_a, \boldsymbol{\eta}_e, \mathbf{X}_n) = p(h_n | a_n, \mathbf{e}_n)\, p(a_n | \boldsymbol{\eta}_a, \mathbf{x}_n^a) \left(\prod_{i=1}^{E_n} p(e_n^i | \boldsymbol{\eta}_e, \mathbf{x}_n^{e_i}) \right)$$

Pereira, F. C., Rodrigues, F., Polisciuc. E. and Ben-Akiva, M. "Why so many people? Explaining non-habitual public transport overcrowding with internet data". IEEE Transactions on Intelligent Transport Systems. 2015

Modeling demand in planned events

. First we need to determine $\boldsymbol{\eta} = [\boldsymbol{\eta_a}, \boldsymbol{\eta_e}]$

. Reminding Bayes' rule (with $\boldsymbol{D} = \{\mathbf{X}_{\boldsymbol{n}}, h_n\}_{n=1}^{N}$)…

likelihood prior

$$p(\boldsymbol{\eta}|\boldsymbol{D}) = \frac{\boldsymbol{p}(\boldsymbol{\eta}, \boldsymbol{D})}{\boldsymbol{p}(\boldsymbol{D})} = \frac{\boldsymbol{p}(\boldsymbol{D}|\boldsymbol{\eta})\boldsymbol{p}(\boldsymbol{\eta})}{\boldsymbol{p}(\boldsymbol{D})}$$

evidence

Posterior

Modeling demand in planned events

. First we need to determine first the "best" values for $\boldsymbol{\eta} = [\boldsymbol{\eta_a}, \boldsymbol{\eta_e}]$

. Reminding Bayes' rule (with $\boldsymbol{D} = \{\mathbf{X}_{\boldsymbol{n}}, h_n\}_{n=1}^{N}$)…

$$p(\boldsymbol{\eta}|\boldsymbol{D}) = \frac{\boldsymbol{p}(\boldsymbol{\eta}, \boldsymbol{D})}{\boldsymbol{p}(\boldsymbol{D})} = \frac{\boldsymbol{p}(\boldsymbol{D}|\boldsymbol{\eta})\boldsymbol{p}(\boldsymbol{\eta})}{\boldsymbol{p}(\boldsymbol{D})}$$

useful consequence:

$$p(\boldsymbol{\eta}|\boldsymbol{D}) \propto \boldsymbol{p}(\boldsymbol{D}|\boldsymbol{\eta})\boldsymbol{p}(\boldsymbol{\eta})$$

Modeling demand in planned events

Assuming a zero-mean Gaussian prior for $\boldsymbol{\eta}$, we can do:

$$p(\boldsymbol{\eta}|\boldsymbol{D}) \propto \prod_{\boldsymbol{n}=1}^{\boldsymbol{N}} \int p(h_n, a_n, \boldsymbol{e}_n | \boldsymbol{\eta}, \mathbf{X}_n) da_n d\boldsymbol{e}_n \; \boldsymbol{p}(\boldsymbol{\eta})$$

Inference tools:

. Approximate inference – Expectation Propagation (EP), Variational Bayes →Infer.NET

. Monte Carlo methods → STAN package

Modeling demand in planned events

Extension:

. Gaussian Processes priors on routine and events components

. Truncated Gaussian constraints

Details in appendix

Experiments with PT arrivals

Data:

. 5 months of Smartcard data from Singapore

. Focus on 2 areas (which have multiple venues)

. Descriptive statistics:

Area	Average daily arrivals $\pm$ std.	Average daily events $\pm$ std.	Maximum daily events	Num. days without events
Stadium	4101 ($\pm$ 925)	0.230 ($\pm$ 0.554)	3	114 (82.014%)
Expo	15027 ($\pm$ 5515)	2.446 ($\pm$ 1.986)	8	23 (16.547%)

Experiments with PT arrivals

Data:

. Scraping and API based

. Summary:

Source	Num. events study areas	Number of categories	Text description size (± std. dev.)	Retrieval type
eventful.com	1221	28	1112.3 (± 1337.1)	API
singaporeexpo.com.sg	58	28	124.9 (± 159.5)	scrapper
last.fm	11	-	901.2 (± 1037.5)	API
timeoutsingapore.com	568	49	411.8 (± 866.6)	scraper

A quick note on text analysis

. Text represented as bag-of-words

. Vector of word frequencies

. Dimensionality of entire dictionary (N)

. Topic modeling

. Dimensionality reduction

. Text re-represented as a linear combination of K<<N vectors (a.k.a.) topics

Generative formulation and examples in appendix

Experiments with PT arrivals

Attributes:

. Routine: time of day; day of week

. Events: topics; category; venue; start/end times; FCBK likes; Google hits; all day/multi-day/short

Different model for each area

Experiments with PT arrivals

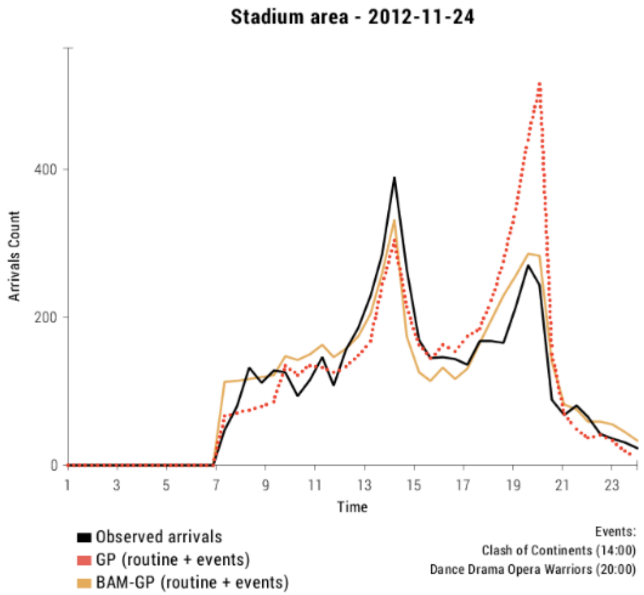
Stadium:

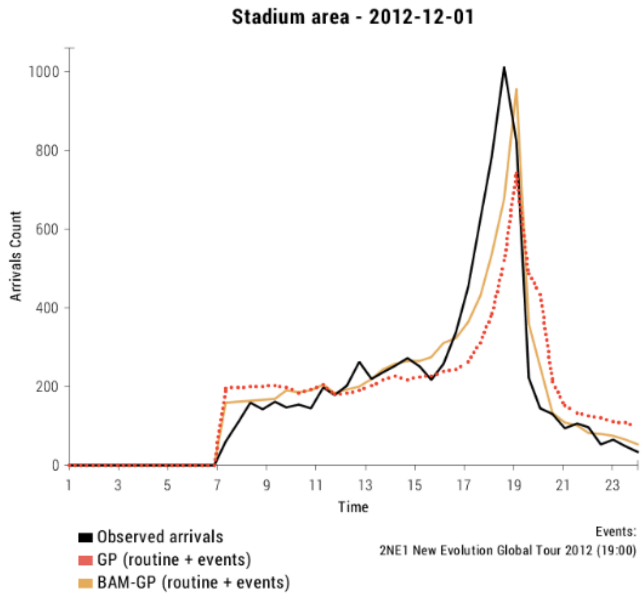
	Evaluation: all times			Evaluation: event periods only		
Model	CorrCoef	RAE	R^2	CorrCoef	RAE	R^2
Linear Reg. (routine only)	0.646	0.642	0.417	0.649	0.735	0.092
Linear Reg. (routine + events)	0.739	0.620	0.543	0.709	0.620	0.502
GP (routine only)	0.667	0.616	0.445	0.654	0.707	0.117
GP (routine + events)	0.777	0.567	0.603	0.751	0.581	0.564
BAM-LR	0.737	0.605	0.544	0.694	0.582	0.474
BAM-GP	**0.795**	**0.556**	**0.632**	**0.811**	**0.503**	**0.658**

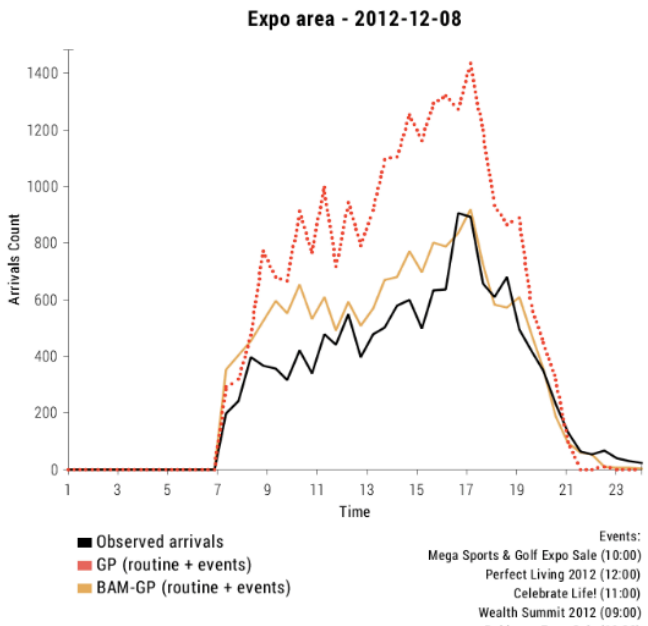
Expo:

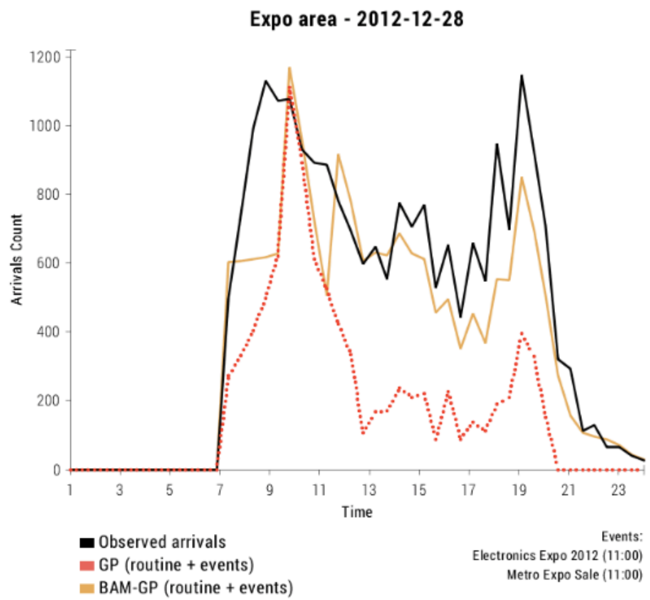
	Evaluation: all times			Evaluation: event periods only		
Model	CorrCoef	RAE	R^2	CorrCoef	RAE	R^2
Linear Reg. (routine only)	0.581	0.723	0.338	0.390	0.816	0.098
Linear Reg. (routine + events)	0.707	0.617	0.500	0.557	0.743	0.300
GP (routine only)	0.718	0.576	0.514	0.621	0.670	0.341
GP (routine + events)	0.750	0.547	0.543	0.676	0.668	0.382
BAM-LR	0.661	0.652	0.436	0.484	0.772	0.229
BAM-GP	**0.796**	**0.472**	**0.633**	**0.736**	**0.565**	**0.540**

Experiments with PT arrivals

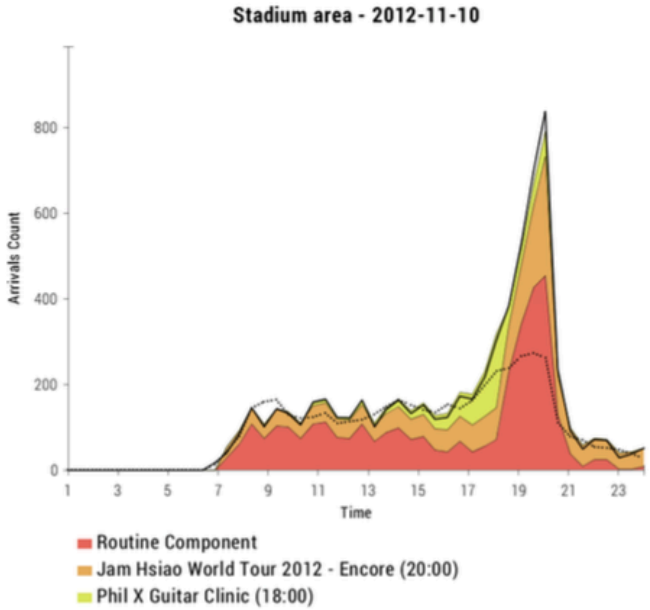




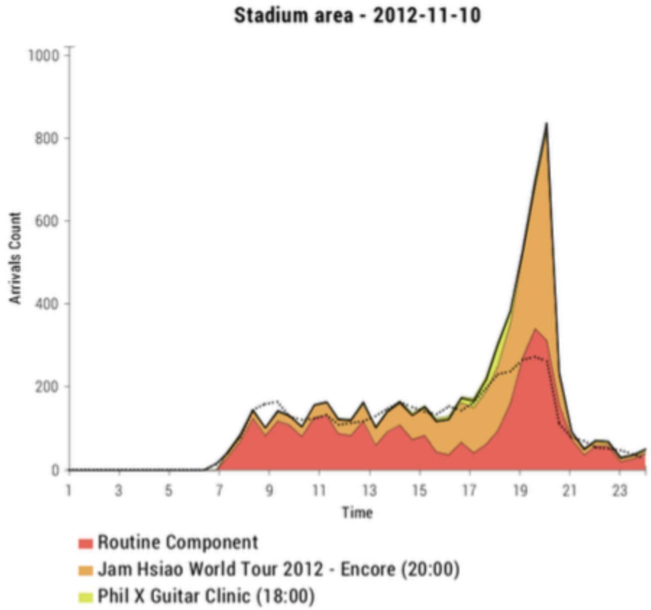




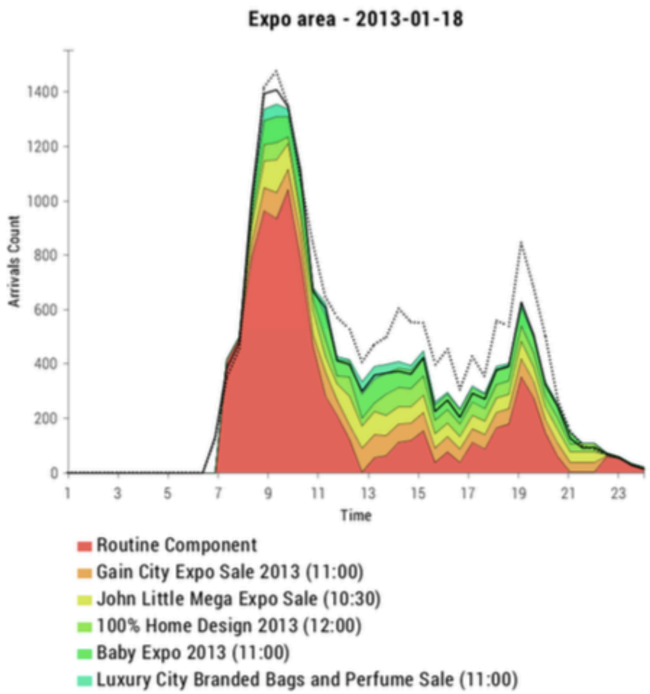
Experiments with PT arrivals

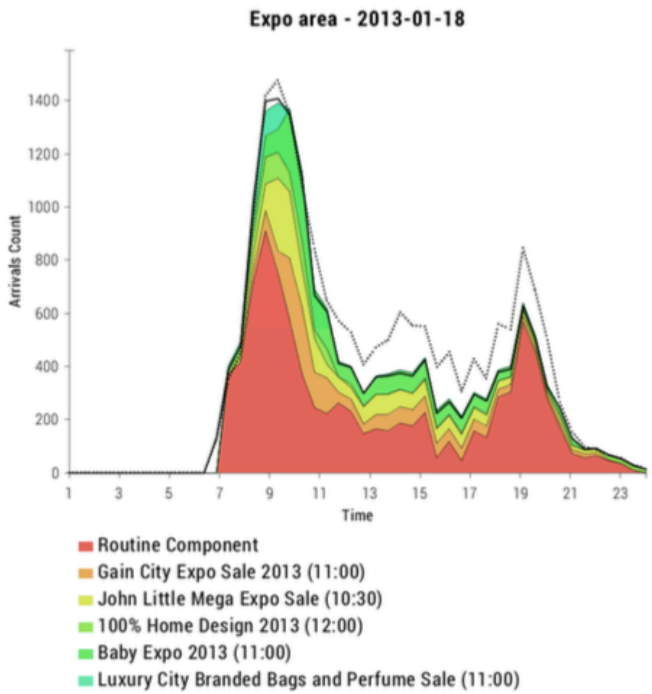


(b) BAM-LR



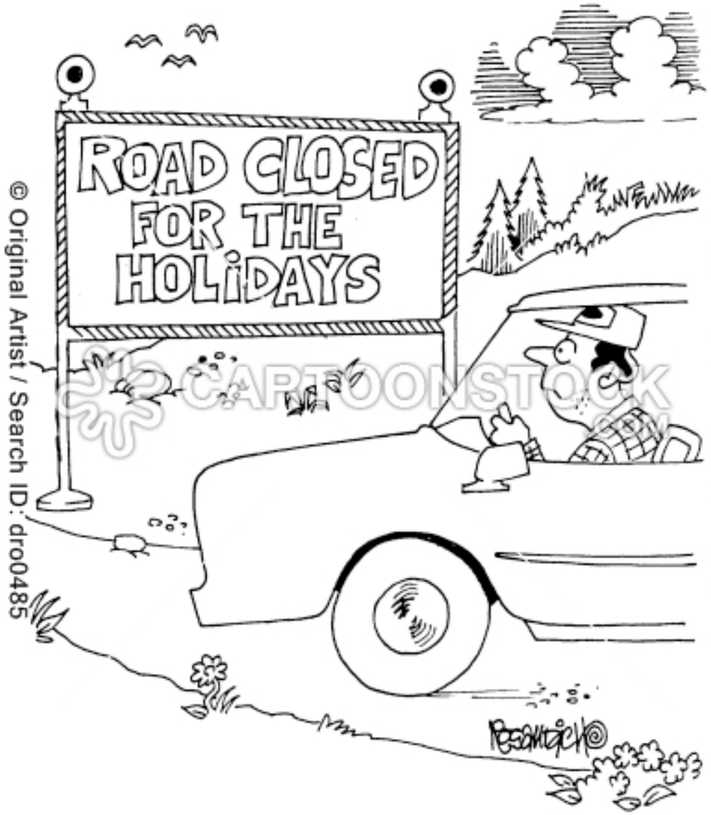
(c) BAM-GP





Predictable supply

www.clker.com



Predictable supply

Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	**Predictable**
crisis situations, unknown high fluctuations	incidents, weather, crisis situations	**Non predictable**

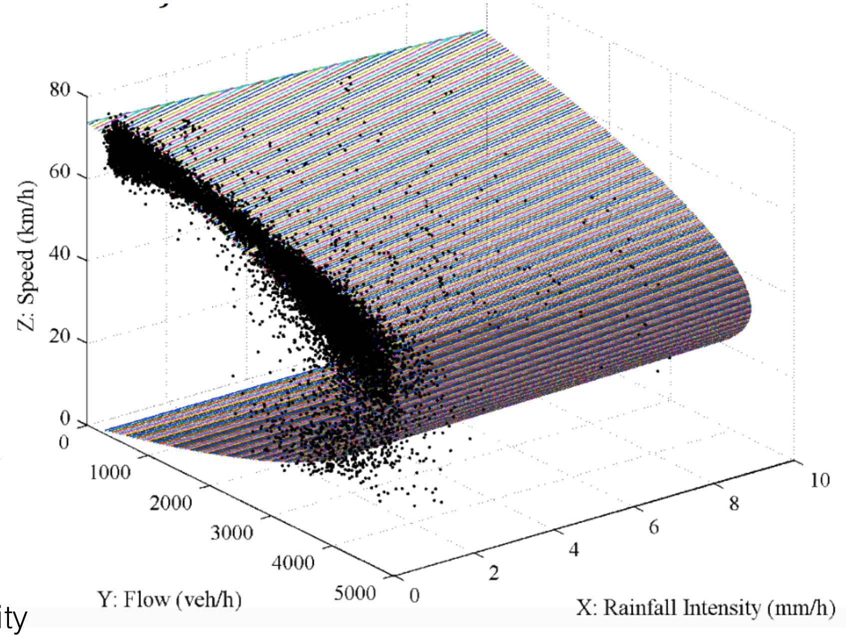
Predictable supply

. Road works information feed

. Planned closures (e.g. national day)

Changes in capacity

→**The trick is to get the data!**



. Weather forecasts

Changes in speed density functions

Observed speed

Free flow speed

Obs density

Min Speed

Jam density

$$\begin{cases} u = u_0 + (u_f - u_0)(1 - \dfrac{k}{k_j})^{\alpha} \\ q = uk_j[1 - (\dfrac{u - u_0}{u_f - u_0})^{1/\alpha}] \end{cases}$$

Observed flow

Rainfall

$$u_f(r) = ar^b + c$$

Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	Predictable
crisis situations, unknown high fluctuations	Incidents, weather, crisis situations	Nonpredictable

Non predictable demand

Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	**Predictable**
crisis situations, unknown high fluctuations	Incidents, weather, crisis situations	**Nonpredictable**

Non predictable demand

Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	**Predictable**
crisis situations, unknown high fluctuations	incidents, weather, crisis situations	**Non predictable**

Non predictable demand

Lack of information

. Where? When? What?

Can real-time feeds help?

. Twitter, Facebook

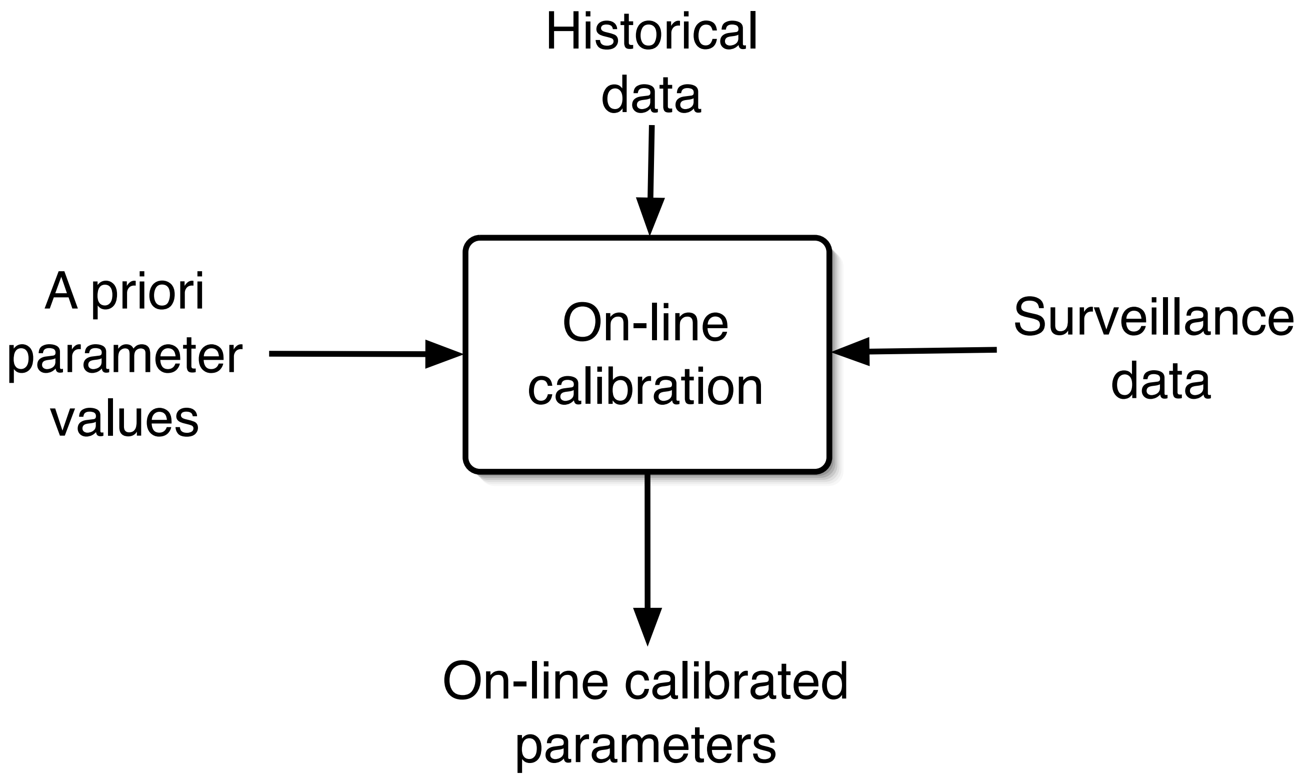
. Radio, TV

. Traffic surveillance

On the semantic side, still an open challenge

Non predictable demand

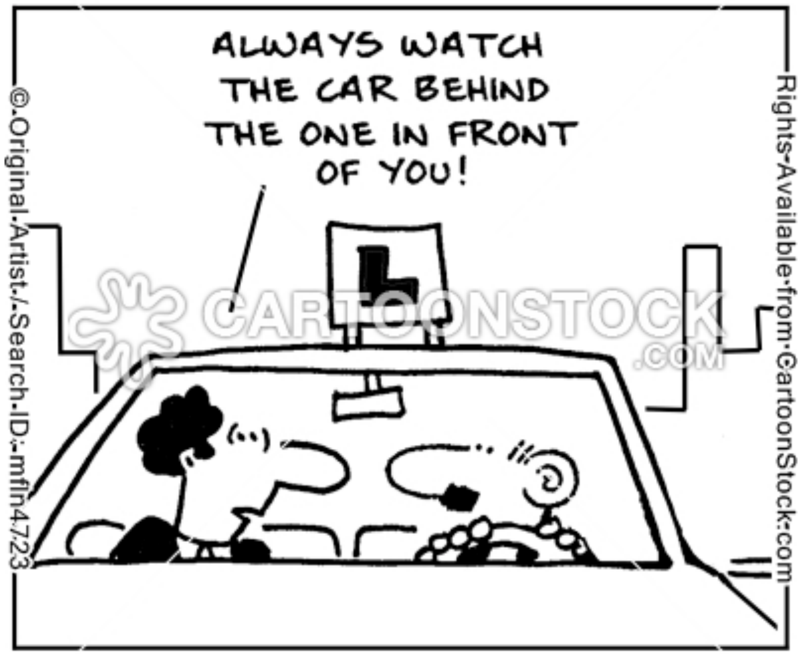
DynaMIT's approach



C. Antoniou, M. Ben-Akiva, and H. N. Koutsopoulos. On-Line Calibration of Traffic Prediction Models. Transportation Research Record: Journal of the Transportation Research Board 1934, pp. 235-245, Washington D.C., 2005.

Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	**Predictable**
crisis situations, unknown high fluctuations	Incidents, weather, crisis situations	**Nonpredictable**

Non predictable supply



Non predictable supply

Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	**Predictable**
crisis situations, unknown high fluctuations	incidents, weather, crisis situations	**Non predictable**

Non predictable supply

Data

. Incident response feed (communications)

. Environmental sensing

. Traffic surveillance feed

. Incident detection

Challenges

. How serious (capacity reduction)?

. How long (incident duration)?

. How do things change (speed-density functions)?

Demand	Supply	
special events, demonstrations holidays	road works, closures, mega events, weather	**Predictable**
crisis situations, unknown high fluctuations	Incidents, weather, crisis situations	**Nonpredictable**

Non predictable supply

Data

. **Incident response feed (communications)**

. Environmental sensing

. Traffic surveillance feed

. **Incident detection**

Challenges

. **How serious (capacity reduction)?**

. **How long (incident duration)?**

. How do things change (speed-density functions)?

Demand	**Supply**	
special events, demonstrations holidays	road works, closures, mega events, weather	**Predictable**
crisis situations, unknown high fluctuations	Incidents, weather, crisis situations	**Nonpredictable**

Data

Archive with 2 years (2010/11) of all incident records from Singapore. For each incident:

. Lanes blocked

. nr vehicles

. Location

. Time

. **communication log**.

Focus on **Singapore expressways** (total of 10139 events)

Data

Example:

id: 473586

zone ID: 2

Location (X, Y): 26266.6, 34916.9

Road name: AYE

Location type: 3

Lane blockage: Lane 1, Shoulder

Down point: 20.32

Congestion status: 0

Start time: 2010-08-20 22:50:01

End time: 2010-08-20 23:31:45

Number of vehicles: 2

Comm. log:

2250hrs - TP Joe X spots an accident. Car and bike involved.

2255hrs - Passers-by shift the bike to the shoulder.

2300hrs - Ambulance arrives at location. LTM arrives at location.

2309hrs - Ambulance conveys rider to National University Hospital.

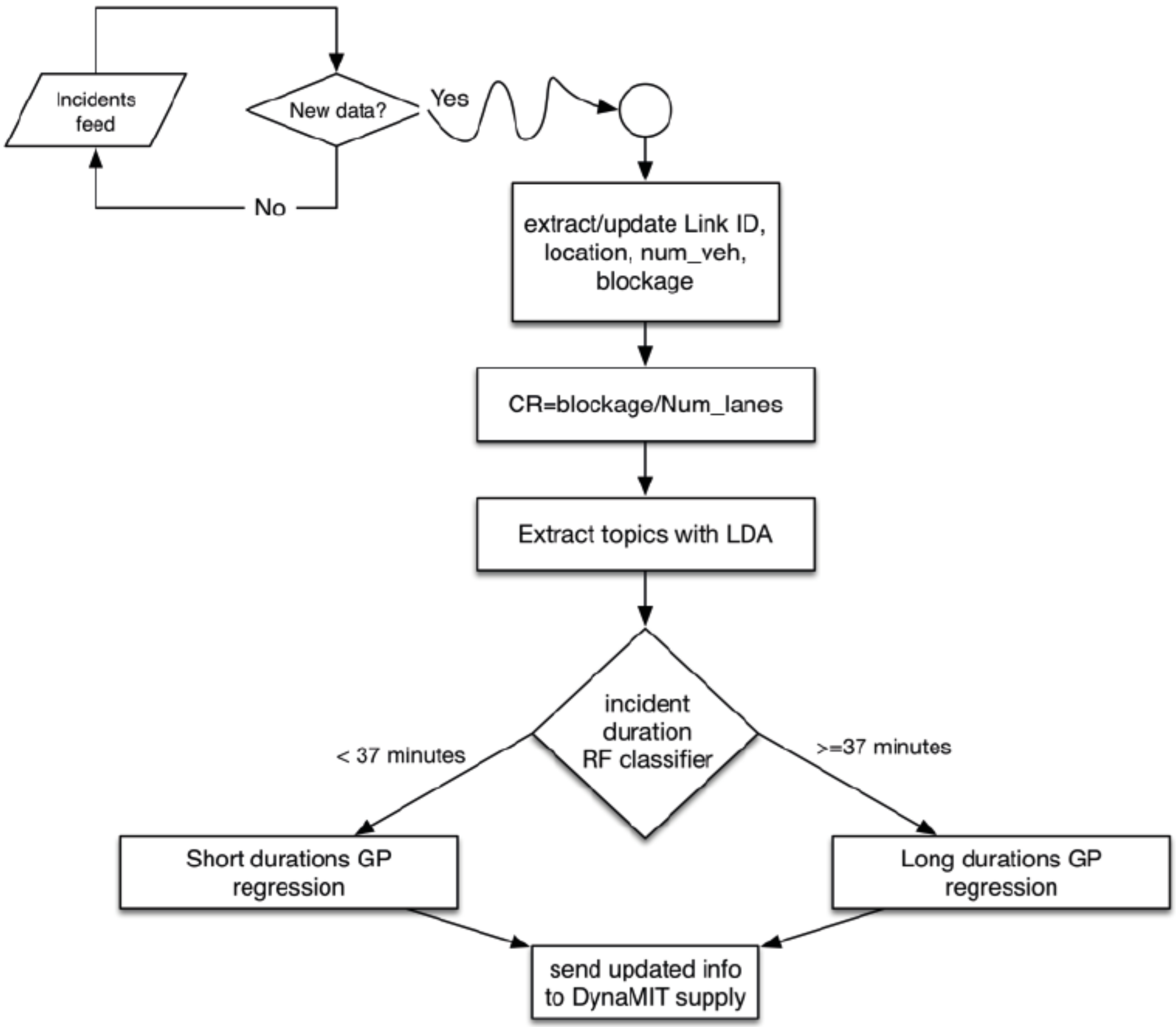
2310hrs - TP arrives at location.

2311hrs - Notify by LTM the rider is seriously injured. The accident involves a car and bike.

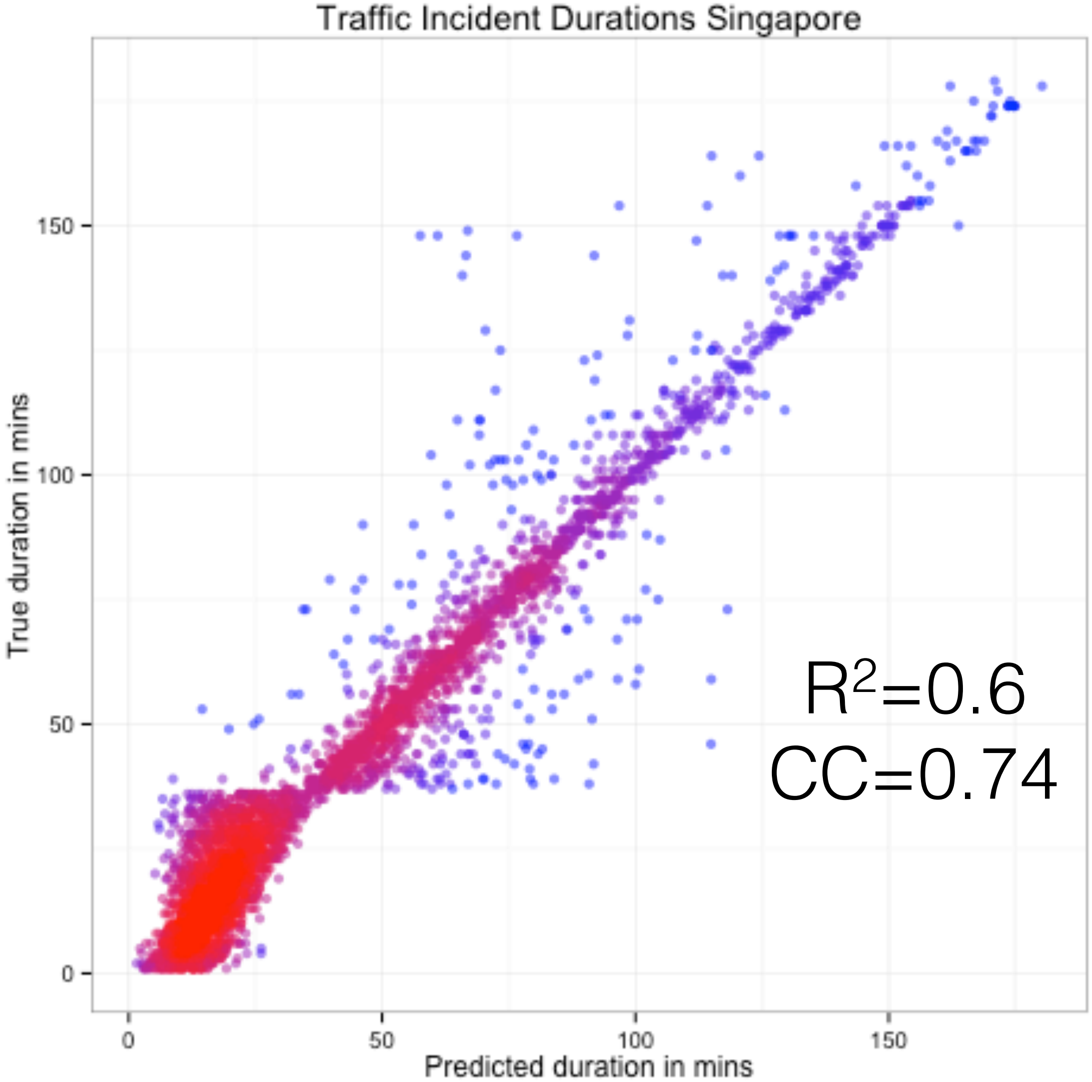
2331hrs - TP requests RC and LTM to resume patrolling. All other vehicles move off. Shoulder clear

Note: This example does not correspond any specific incident in the database, but mimics the type of information available

Methodology



Results



R^2=0.6
CC=0.74

Results

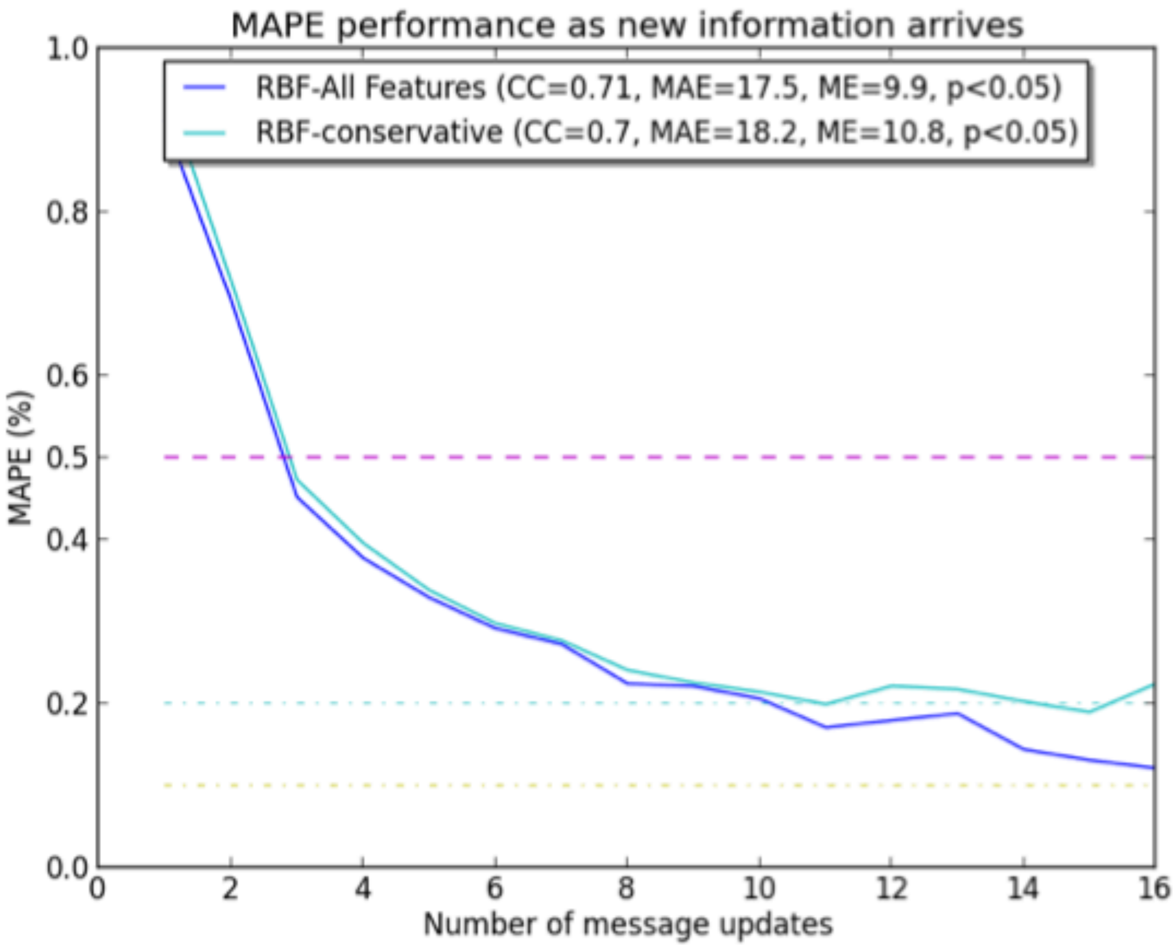
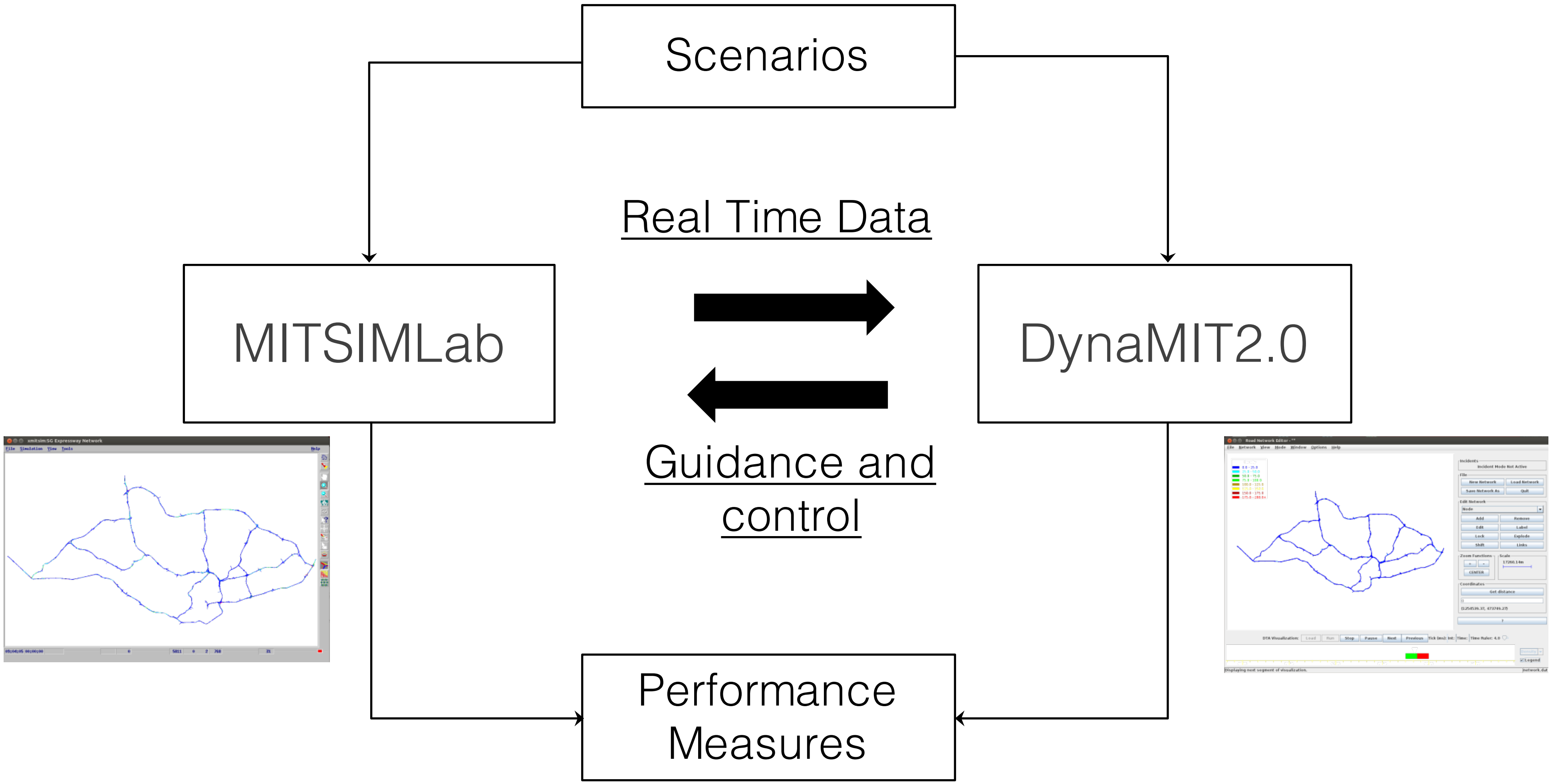


Figure 6: Mean Absolute Percentage Error with number of messages.

Pereira, F. C., Rodrigues, F. and Ben-Akiva, M. ”Text analysis in incident duration prediction”. Transport Research Part C. Volume 37, Pages 177-192, Elsevier, 2013.

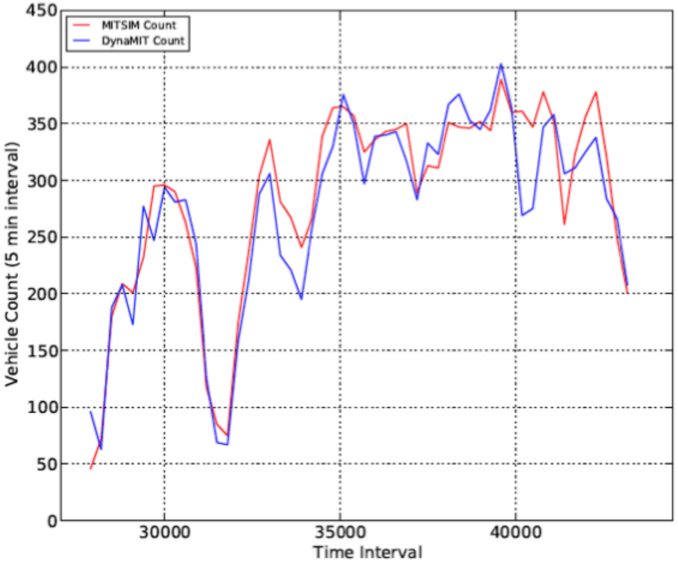
Impact in DynaMIT

- Numerical experiments with closed-loop framework (**CLF**), DynaMIT2.0 with a traffic microsimulator

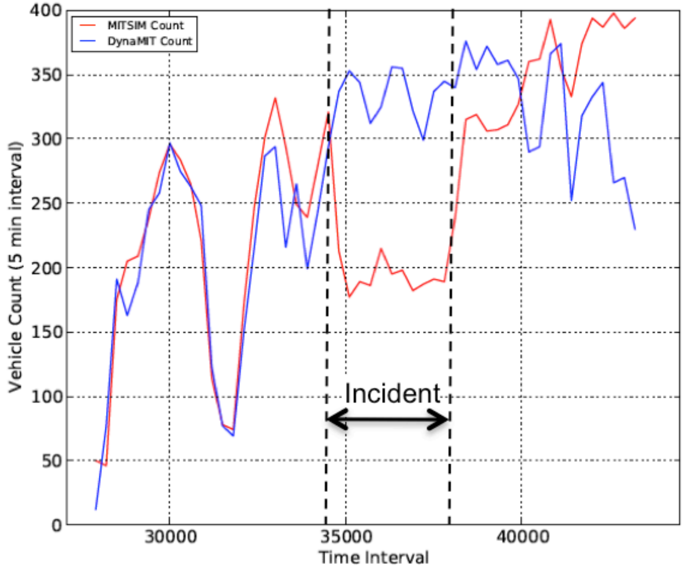


Impact in DynaMIT

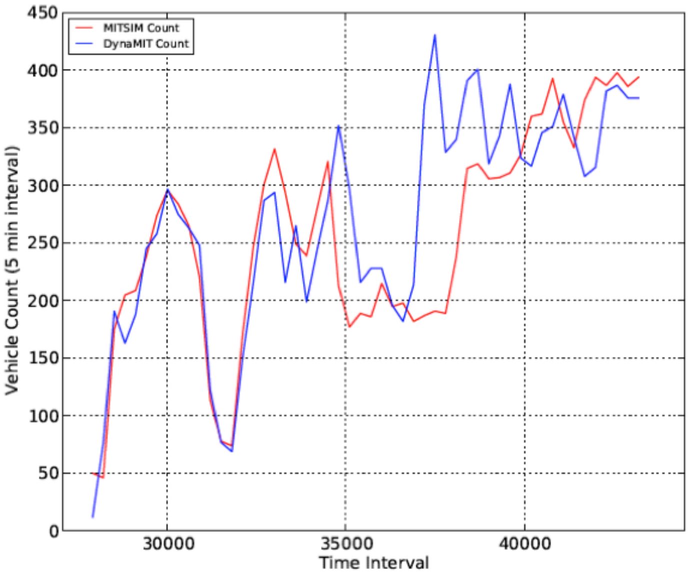
Sensor Counts: Incident affected segment



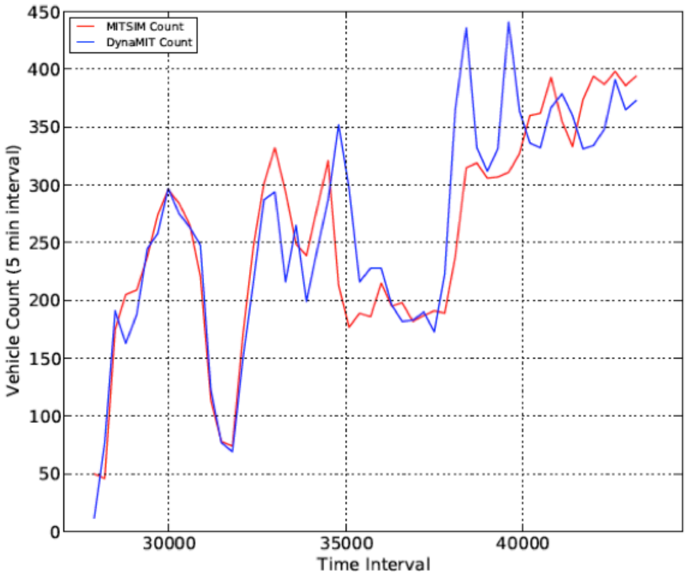
(a) MISTIM vs DynaMIT (Base - No Incident)



(b) MITSIM vs DynaMIT with no IF



(c) MITSIM vs DynaMIT with inferred IF from SAM



(d) MITSIM vs DynaMIT with true IF

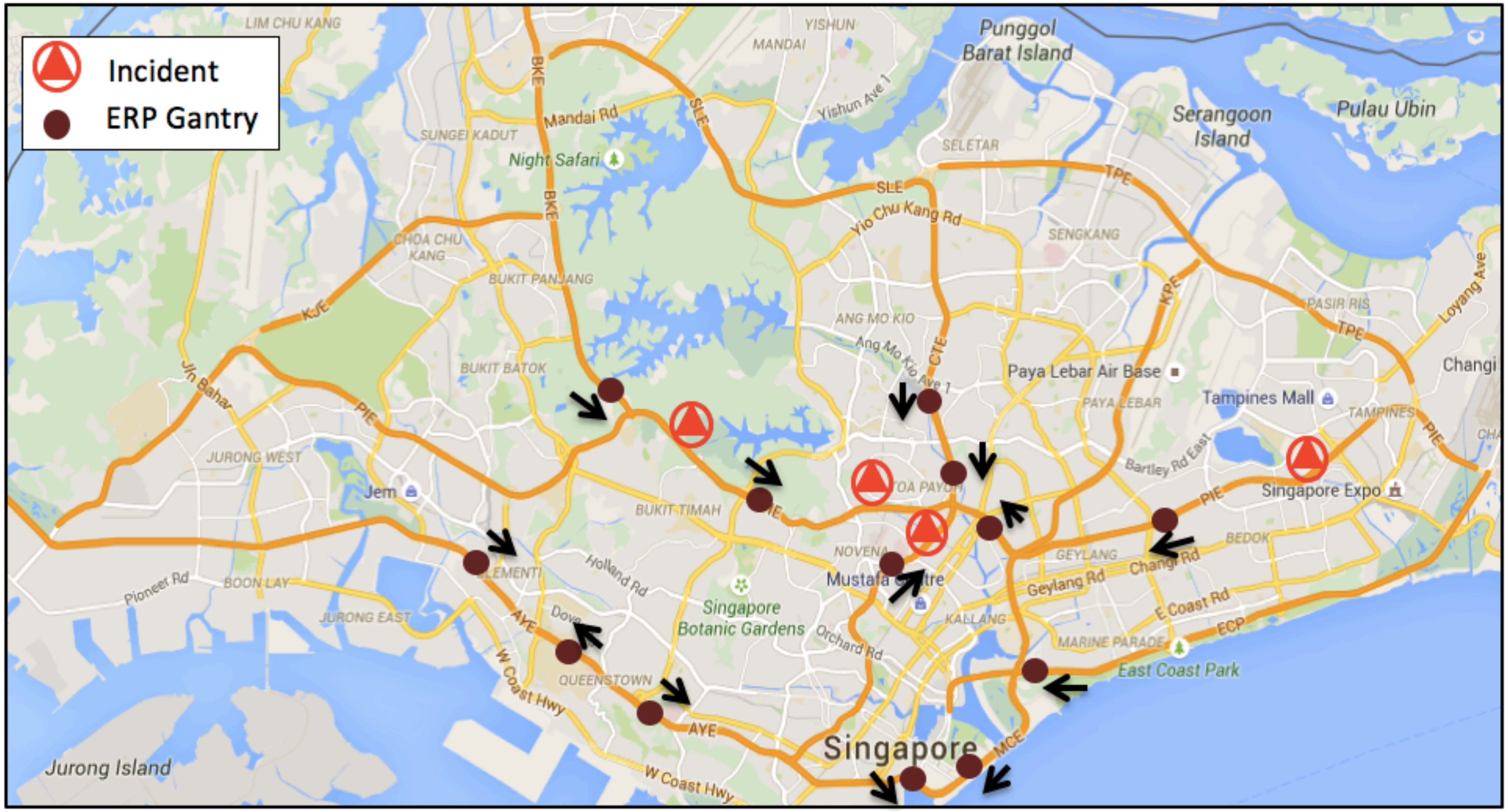
Error Comparison: only affected segments

Scenario	RMSN
Without II	56%
Predicted II	36%
True II	24.6%

II=Incident Information
RMSN=Root Mean Squared error Normalized

Predictive traffic management: Case Study

Case Study



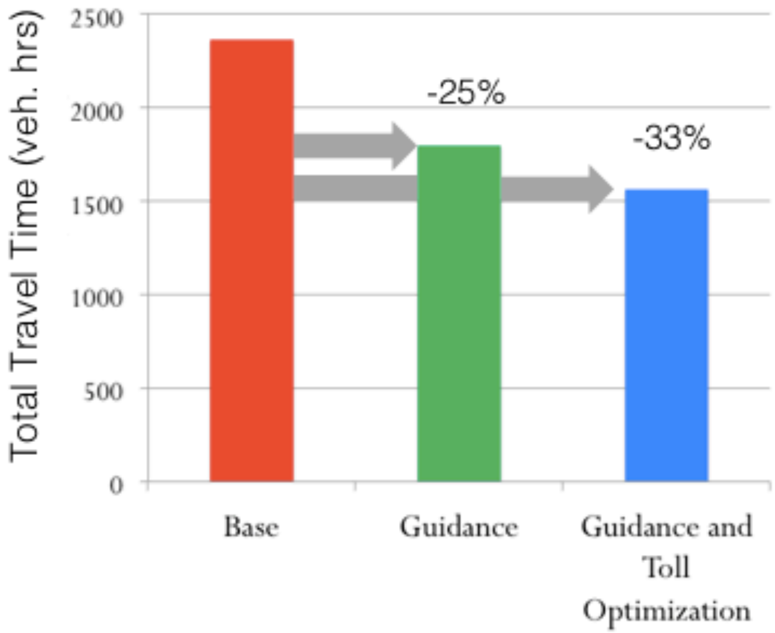
- Incidents/road works from September 15th, 2011
  - Two Singapore expressways, 13 ERP gantries

- Simulation of morning peak traffic: 7:30 AM ~ 2:30 PM
  - Toll rates changing at 5 min interval

Three scenarios

Base case	No guidance
Guidance	Predictive guidance
Guidance and toll optimization	guidance and optimized tolls

Reduction in Network Delay



Scenario	Delay of affected drivers** (veh. hrs)	Total delay (veh. hrs)
Base	2184	87645
Guidance	1648 (-25%)	81626 (-7%)
Guidance & toll optimization	1473 (-33%)	79141 (-10%)

***Affected vehicles are defined as vehicles passing incident locations*

Wrapping up

Conclusions

Non-recurrent events challenging for transport prediction

. Scarce data, difficult to use

. Statistical methods: too much reliance on historical patterns

. Model-based approach: not efficient enough in real-time

Proposed approach:

. Explore wide range data sources (including textual)

. Combine statistical with model-based techniques

Collaborations: MIT, Singapore (SMART), Filipe Rodrigues, Stanislav Borysov, Haizheng Zhang, Constantinos Antoniou, Moshe Ben-Akiva

Appendix

On topic modeling

- Topic modeling – Latent Dirichlet Allocation (LDA) (Blei et al 2012)

On topic modeling

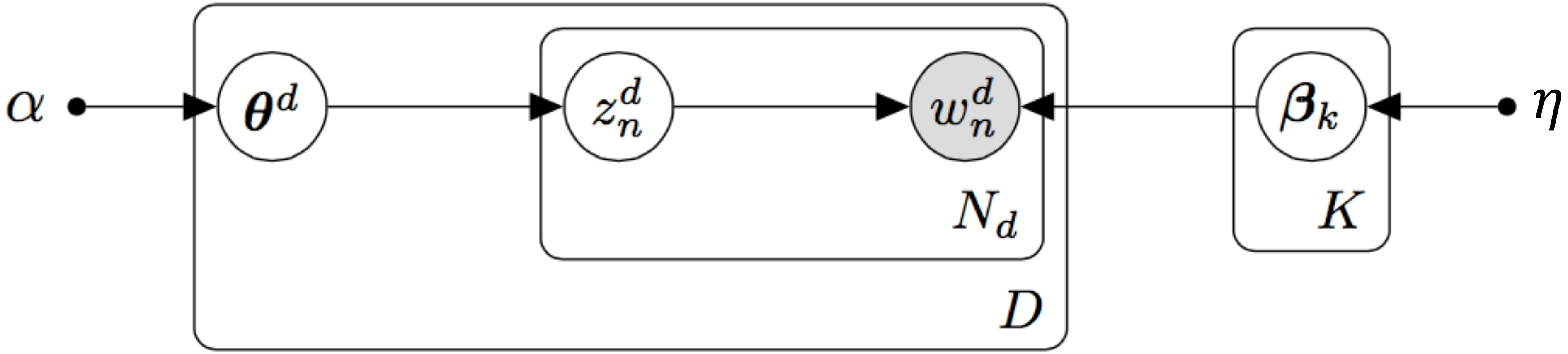
- Topic modeling – Latent Dirichlet Allocation (LDA) (Blei et al 2012)
- Based on assumptions on probabilistic word distributions by topics
  - Each document is distribution of topics, each topic is a distribution of words

On topic modeling

- Topic modeling – Latent Dirichlet Allocation (LDA) (Blei et al 2012)
- Based on assumptions on probabilistic word distributions by topics
  - Each document is distribution of topics, each topic is a distribution of words
- Generative algorithm (K=number of topics, *α* and *η* are hyperparameters for Dirichlet priors)

1) Draw a topic β_k from $\beta_k \sim$ Dirichlet(η) for $k = 1 \ldots K$
2) For each document d:
   a) Draw topics proportions θ_d such that $\theta_d \sim$ Dirichlet(α)
   b) For each word $w_{d,n}$:
      i) Draw topic assignment $z_{d,n} \sim$ Multinomial(θ_d)
      ii) Draw word $w_{d,n} \sim$ Multinomial($\beta_{z_{d,n}}$)

LDA graphical model



On topic modeling

- The previous algorithm translates into the following:

$$p(\theta,\beta \mid \mathbf{w},\alpha,\eta) = \frac{p(\mathbf{w} \mid \theta,\beta,\alpha,\eta)p(\theta,\beta \mid \alpha,\eta)}{p(\mathbf{w} \mid \alpha,\eta)}$$

with

$$p(\mathbf{w} \mid \theta,\beta,\alpha,\eta) = \sum_{k=1}^{K}\prod_{d=1}^{D}\prod_{n=1}^{N} p(z=k \mid \theta_d)p(w_{d,n} \mid \beta_z)$$

$$p(\theta,\beta \mid \alpha,\eta) = \prod_{k=1}^{K} p(\beta_k \mid \eta)\prod_{d=1}^{D} p(\theta_d \mid \alpha)$$

- Only $w_{d,n}$ are observed, so **z** are latent variables
- In general, direct inference is intractable. The common solutions are Expectation-Maximization (EM) or Gibs sampling

SMART

Example with incidents

- An example with K=6, α and η=1/6 with our dataset that has 10139 with 3545 different words overall

- tp #0: 0.15*noinjur + 0.09*veh + 0.08*noinfst + 0.07*came + 0.06*across + 0.05*selfdriven
- tp #1: 0.11*open + 0.09*spill + 0.09*oil + 0.09*scdf + 0.08*close + 0.06*2ln
- tp #2: 0.06*polic + 0.06*drive + 0.04*crew + 0.04*drink + 0.03*div + 0.03*driver
- tp #3: 0.07*ab + 0.07*tow + 0.06*convei + 0.06*rider + 0.06*hosp + 0.05*bike
- tp #4: 0.08*damag + 0.07*call + 0.04*tow + 0.04*veh + 0.03*tp + 0.03*vig
- tp #5: 0.06*tp + 0.05*convei + 0.05*ab + 0.04*tow + 0.04*hosp + 0.04*veh

Our earlier example would have the following assignment (0.01, 0.02, 0.02, 0.32, 0.13, 0.50)

- **Note 1:** we show **only** top 6 words of each topic
- **Note 2:** we skipped all the details in data preparation and cleaning!

Examples from events

Some topics

Live concerts, dance events, DJs…

topic #0: 0.163*exhibit + 0.103*film + 0.022*solo + 0.021*la + 0.019*best + 0.019*problem
topic #2: 0.059*expo + 0.055*train + 0.029*bar + 0.023*dinner + 0.023*youth + 0.022*wine
topic #3: 0.037*run + 0.032*littl + 0.031*2nd + 0.029*natur + 0.025*trail + 0.024*india + 0.022*sentosa
topic #4: 0.115*intern + 0.088*confer + 0.057*book + 0.053*nation + 0.049*annual + 0.037*food
topic #7: 0.096*parti + 0.085*concert + 0.040*christma + 0.039*launch + 0.038*garden + 0.033*xma
topic #9: 0.151*live + 0.062*danc + 0.038*life + 0.029*collect + 0.028*esplanad + 0.021*present
topic #13: 0.095*world + 0.054*present + 0.039*session + 0.034*zouk + 0.033*vs + 0.031*ft + 0.026*dj
topic #10: 0.056*storytel + 0.054*chines + 0.047*english + 0.040*friend + 0.024*public+ 0.023*season
topic #11: 0.059*seminar + 0.041*open + 0.040*asian + 0.032*celebr + 0.026*hous + 0.022*properti
topic #17: 0.058*trade + 0.033*manag + 0.031*onlin + 0.025*stream + 0.024*theatr + 0.023*preview
topic #15: 0.106*night + 0.060*museum + 0.031*famili + 0.022*chroam + 0.021*happi + 0.019*arm
itopic #22: 0.213*art + 0.036*game + 0.028*certif + 0.028*footbal + 0.024*school + 0.019*rock
topic #19: 0.052*design + 0.047*forum + 0.036*park + 0.032*stock + 0.031*paint + 0.019*jurong
topic #20: 0.048*talk + 0.033*movi + 0.033*screen + 0.031*nu + 0.029*competit + 0.028*social
topic #21: 0.169*sale + 0.029*warehous + 0.025*librari + 0.022*bag + 0.022*exam + 0.021*mr25
topic #23: 0.270*festiv + 0.050*stori + 0.037*photographi + 0.030*tv + 0.029*introduct + 0.019*write
topic #24: 0.167*asia + 0.043*invest + 0.040*meet + 0.032*anniversari + 0.025*comedi + 0.017*profession

Examples from events

Some topics

Cultural events, arts, theatre, cinema

topic #0: 0.163*exhibit + 0.103*film + 0.022*solo + 0.021*la + 0.019*best + 0.019*problem
topic #2: 0.059*expo + 0.055*train + 0.029*bar + 0.023*dinner + 0.023*youth + 0.022*wine
topic #3: 0.037*run + 0.032*littl + 0.031*2nd + 0.029*natur + 0.025*trail + 0.024*india + 0.022*sentosa
topic #4: 0.115*intern + 0.088*confer + 0.057*book + 0.053*nation + 0.049*annual + 0.037*food
topic #7: 0.096*parti + 0.085*concert + 0.040*christma + 0.039*launch + 0.038*garden + 0.033*xma
topic #9: 0.151*live + 0.062*danc + 0.038*life + 0.029*collect + 0.028*esplanad + 0.021*present
topic #13: 0.095*world + 0.054*present + 0.039*session + 0.034*zouk + 0.033*vs + 0.031*ft + 0.026*dj
topic #10: 0.056*storytel + 0.054*chines + 0.047*english + 0.040*friend + 0.024*public+ 0.023*season
topic #11: 0.059*seminar + 0.041*open + 0.040*asian + 0.032*celebr + 0.026*hous + 0.022*properti
topic #17: 0.058*trade + 0.033*manag + 0.031*onlin + 0.025*stream + 0.024*theatr + 0.023*preview
topic #15: 0.106*night + 0.060*museum + 0.031*famili + 0.022*chroam + 0.021*happi + 0.019*arm
itopic #22: 0.213*art + 0.036*game + 0.028*certif + 0.028*footbal + 0.024*school + 0.019*rock
topic #19: 0.052*design + 0.047*forum + 0.036*park + 0.032*stock + 0.031*paint + 0.019*jurong
topic #20: 0.048*talk + 0.033*movi + 0.033*screen + 0.031*nu + 0.029*competit + 0.028*social
topic #21: 0.169*sale + 0.029*warehous + 0.025*librari + 0.022*bag + 0.022*exam + 0.021*mr25
topic #23: 0.270*festiv + 0.050*stori + 0.037*photographi + 0.030*tv + 0.029*introduct + 0.019*write
topic #24: 0.167*asia + 0.043*invest + 0.040*meet + 0.032*anniversari + 0.025*comedi + 0.017*profession

Examples from events

Some topics

Outdoors events, sports

topic #0: 0.163*exhibit + 0.103*film + 0.022*solo + 0.021*la + 0.019*best + 0.019*problem
topic #2: 0.059*expo + 0.055*train + 0.029*bar + 0.023*dinner + 0.023*youth + 0.022*wine
topic #3: 0.037*run + 0.032*littl + 0.031*2nd + 0.029*natur + 0.025*trail + 0.024*india + 0.022*sentosa
topic #4: 0.115*intern + 0.088*confer + 0.057*book + 0.053*nation + 0.049*annual + 0.037*food
topic #7: 0.096*parti + 0.085*concert + 0.040*christma + 0.039*launch + 0.038*garden + 0.033*xma
topic #9: 0.151*live + 0.062*danc + 0.038*life + 0.029*collect + 0.028*esplanad + 0.021*present
topic #13: 0.095*world + 0.054*present + 0.039*session + 0.034*zouk + 0.033*vs + 0.031*ft + 0.026*dj
topic #10: 0.056*storytel + 0.054*chines + 0.047*english + 0.040*friend + 0.024*public+ 0.023*season
topic #11: 0.059*seminar + 0.041*open + 0.040*asian + 0.032*celebr + 0.026*hous + 0.022*properti
topic #17: 0.058*trade + 0.033*manag + 0.031*onlin + 0.025*stream + 0.024*theatr + 0.023*preview
topic #15: 0.106*night + 0.060*museum + 0.031*famili + 0.022*chroam + 0.021*happi + 0.019*arm
itopic #22: 0.213*art + 0.036*game + 0.028*certif + 0.028*footbal + 0.024*school + 0.019*rock
topic #19: 0.052*design + 0.047*forum + 0.036*park + 0.032*stock + 0.031*paint + 0.019*jurong
topic #20: 0.048*talk + 0.033*movi + 0.033*screen + 0.031*nu + 0.029*competit + 0.028*social
topic #21: 0.169*sale + 0.029*warehous + 0.025*librari + 0.022*bag + 0.022*exam + 0.021*mr25
topic #23: 0.270*festiv + 0.050*stori + 0.037*photographi + 0.030*tv + 0.029*introduct + 0.019*write
topic #24: 0.167*asia + 0.043*invest + 0.040*meet + 0.032*anniversari + 0.025*comedi + 0.017*profession

Examples from events

Some topics

Festivals, exhibitions, sales

topic #0: 0.163*exhibit + 0.103*film + 0.022*solo + 0.021*la + 0.019*best + 0.019*problem
topic #2: 0.059*expo + 0.055*train + 0.029*bar + 0.023*dinner + 0.023*youth + 0.022*wine
topic #3: 0.037*run + 0.032*littl + 0.031*2nd + 0.029*natur + 0.025*trail + 0.024*india + 0.022*sentosa
topic #4: 0.115*intern + 0.088*confer + 0.057*book + 0.053*nation + 0.049*annual + 0.037*food
topic #7: 0.096*parti + 0.085*concert + 0.040*christma + 0.039*launch + 0.038*garden + 0.033*xma
topic #9: 0.151*live + 0.062*danc + 0.038*life + 0.029*collect + 0.028*esplanad + 0.021*present
topic #13: 0.095*world + 0.054*present + 0.039*session + 0.034*zouk + 0.033*vs + 0.031*ft + 0.026*dj
topic #10: 0.056*storytel + 0.054*chines + 0.047*english + 0.040*friend + 0.024*public+ 0.023*season
topic #11: 0.059*seminar + 0.041*open + 0.040*asian + 0.032*celebr + 0.026*hous + 0.022*properti
topic #17: 0.058*trade + 0.033*manag + 0.031*onlin + 0.025*stream + 0.024*theatr + 0.023*preview
topic #15: 0.106*night + 0.060*museum + 0.031*famili + 0.022*chroam + 0.021*happi + 0.019*arm
itopic #22: 0.213*art + 0.036*game + 0.028*certif + 0.028*footbal + 0.024*school + 0.019*rock
topic #19: 0.052*design + 0.047*forum + 0.036*park + 0.032*stock + 0.031*paint + 0.019*jurong
topic #20: 0.048*talk + 0.033*movi + 0.033*screen + 0.031*nu + 0.029*competit + 0.028*social
topic #21: 0.169*sale + 0.029*warehous + 0.025*librari + 0.022*bag + 0.022*exam + 0.021*mr25
topic #23: 0.270*festiv + 0.050*stori + 0.037*photographi + 0.030*tv + 0.029*introduct + 0.019*write
topic #24: 0.167*asia + 0.043*invest + 0.040*meet + 0.032*anniversari + 0.025*comedi + 0.017*profession

Examples from events

Some topics

Seminars, conferences, storytelling

topic #0: 0.163*exhibit + 0.103*film + 0.022*solo + 0.021*la + 0.019*best + 0.019*problem
topic #2: 0.059*expo + 0.055*train + 0.029*bar + 0.023*dinner + 0.023*youth + 0.022*wine
topic #3: 0.037*run + 0.032*littl + 0.031*2nd + 0.029*natur + 0.025*trail + 0.024*india + 0.022*sentosa
topic #4: 0.115*intern + 0.088*confer + 0.057*book + 0.053*nation + 0.049*annual + 0.037*food
topic #7: 0.096*parti + 0.085*concert + 0.040*christma + 0.039*launch + 0.038*garden + 0.033*xma
topic #9: 0.151*live + 0.062*danc + 0.038*life + 0.029*collect + 0.028*esplanad + 0.021*present
topic #13: 0.095*world + 0.054*present + 0.039*session + 0.034*zouk + 0.033*vs + 0.031*ft + 0.026*dj
topic #10: 0.056*storytel + 0.054*chines + 0.047*english + 0.040*friend + 0.024*public+ 0.023*season
topic #11: 0.059*seminar + 0.041*open + 0.040*asian + 0.032*celebr + 0.026*hous + 0.022*properti
topic #17: 0.058*trade + 0.033*manag + 0.031*onlin + 0.025*stream + 0.024*theatr + 0.023*preview
topic #15: 0.106*night + 0.060*museum + 0.031*famili + 0.022*chroam + 0.021*happi + 0.019*arm
itopic #22: 0.213*art + 0.036*game + 0.028*certif + 0.028*footbal + 0.024*school + 0.019*rock
topic #19: 0.052*design + 0.047*forum + 0.036*park + 0.032*stock + 0.031*paint + 0.019*jurong
topic #20: 0.048*talk + 0.033*movi + 0.033*screen + 0.031*nu + 0.029*competit + 0.028*social
topic #21: 0.169*sale + 0.029*warehous + 0.025*librari + 0.022*bag + 0.022*exam + 0.021*mr25
topic #23: 0.270*festiv + 0.050*stori + 0.037*photographi + 0.030*tv + 0.029*introduct + 0.019*write
topic #24: 0.167*asia + 0.043*invest + 0.040*meet + 0.032*anniversari + 0.025*comedi + 0.017*profession

Extension to Gaussian Processes

$$y = f_r(\mathbf{x}^r) + \sum_{i=1}^{E} f_e(\mathbf{x}^{e_i}) + \epsilon$$

$$y^r \sim \mathbb{I}(y^r > 0)\,\mathcal{N}(y^r | f_r(\mathbf{x}^r), \beta_r),$$
$$y^{e_i} \sim \mathbb{I}(y^{e_i} > 0)\,\mathcal{N}(y^{e_i} | f_e(\mathbf{x}^{e_i}), \beta_e)$$

Defining $\mathbf{f}$ as the vector with function $f(\mathbf{x})$ evaluated on all vectors $\mathbf{x}$

$$\mathbf{f}^r \sim \mathcal{GP}(m_r(\mathbf{x}^r) \equiv 0, k_r(\mathbf{x}^r, \mathbf{x}^{r'}))$$

$$\mathbf{f}^e \sim \mathcal{GP}(m_e(\mathbf{x}^e) \equiv 0, k_e(\mathbf{x}^e, \mathbf{x}^{e'}))$$

Extension to Gaussian Processes

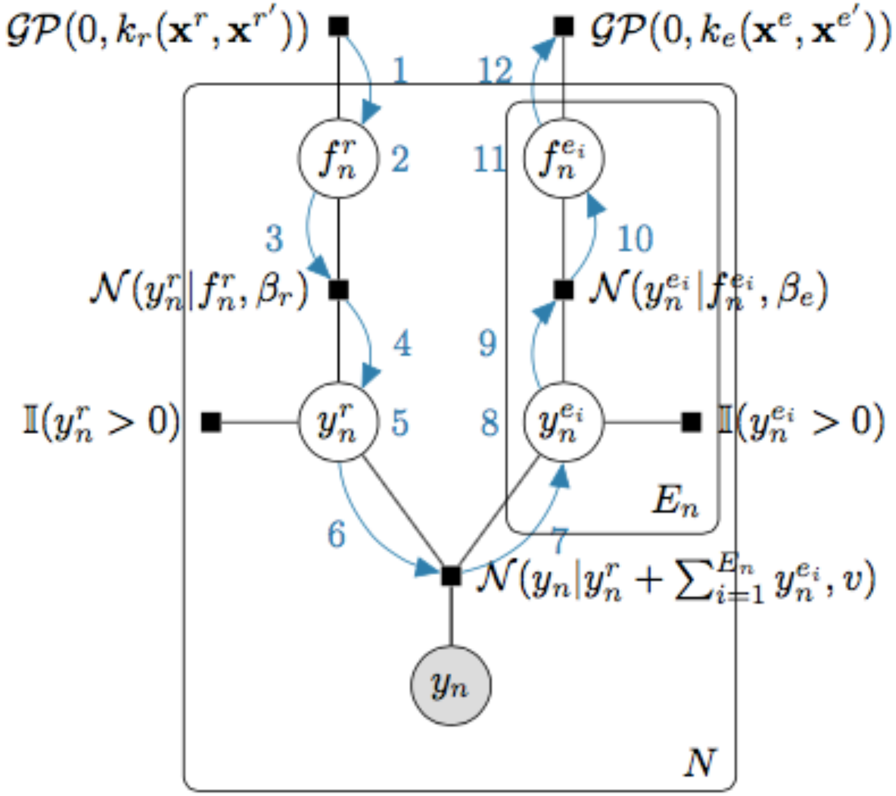
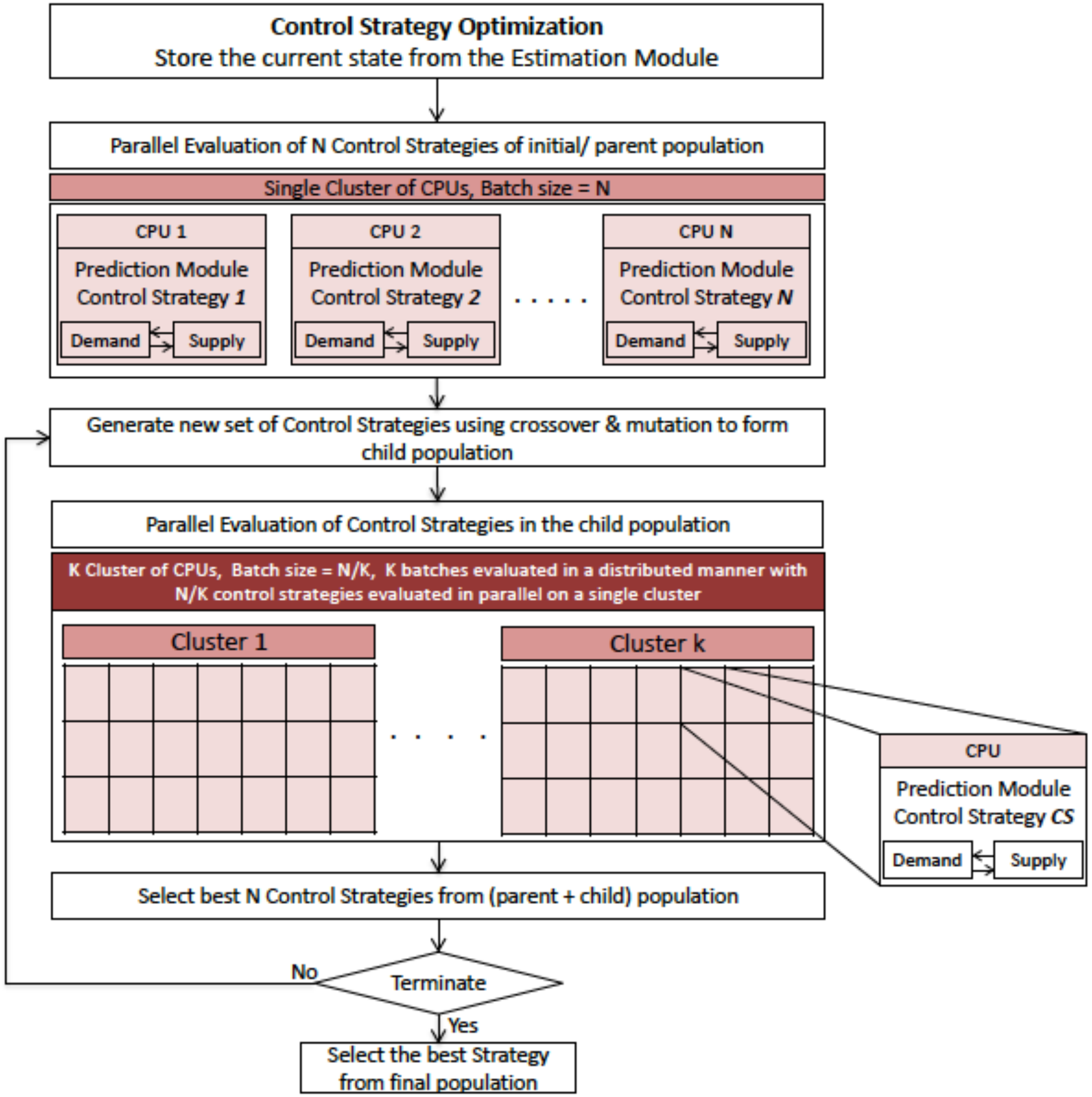


Figure 7.2: Factor graph of the proposed Bayesian additive model with Gaussian process components. The outer plate represents the observations, while the inner plate represents the events associated with each observation. The blue arrows represent the message-passing algorithm for performing approximate Bayesian inference. The second flow of messages starting from the GP factor for the events component that goes in the opposite direction is not shown.

GA based solution approach



Numerical Experiments using DynaMIT2.0

- Application to the problem of real time toll optimization with the objective of minimizing total network travel times

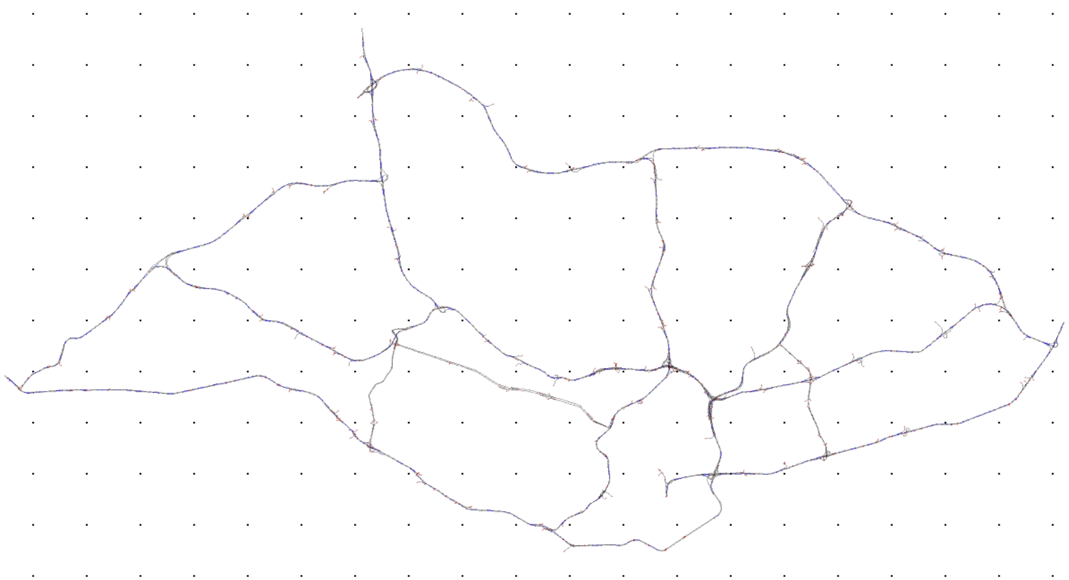
OBJECTIVES

- Examine the performance of the parallel genetic algorithm on the Singapore expressway network

- Examine the impact of providing predictive guidance and optimal tolls in mitigating network congestion under incident scenarios

Performance of Parallel GA

- Singapore Expressway network: 948 nodes, 1150 links, 3891 segments, 4121 OD pairs, 16 toll gantries



- Period between 7:30 AM and 8:30 PM simulated in DynaMIT2.0 using a 5 minute state estimation interval and 15 minute prediction/optimization horizon
- GA parameters - Population size: 60 | cross-over probability: 0.7 | mutation probability: 0.05 | Number of generations: 30 | 25 replications with different initial populations

Performance of Parallel GA

Maximum Toll (SGD)	Travel Time * with static toll	Travel Time with Toll optimization	Percentage Improvement
2.5	480	475	1.00
3	480	473	1.32
3.5	480	472	1.60
4	480	466	2.70
4.5	480	466	2.70

*seconds/vehicle

- Significant travel time savings of up to **2.7%** even under non-recurrent scenarios

Some error formulas

Mean absolute percentage error ("text analysis paper")

$$\mathrm{MAPE} = \frac{1}{|M(t)|} \sum_{i \in M(t)} \left| \frac{\hat{x}_i - x_i}{x_i} \right| \times 100\%$$

where $M(t)$ corresponds to the set of measurements that occurred in time t and x_i and $\hat{x}_i$ the observed and predicted values, respectively. Figure 2 shows

Mean absolute error (Wikipedia)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n} |e_i|.$$

As the name suggests, the mean absolute error is an average of the absolute errors $e_i = |f_i - y_i|$, where f_i is the prediction and y_i the true value. Note that alternative formulations may include relative frequencies as weight factors.

Correlation Coefficient (Wikipedia)

Pearson's correlation coefficient when applied to a sample is commonly represented by the letter *r* and may be referred to as the *sample correlation coefficient* or the *sample Pearson correlation coefficient*. We can obtain a formula for *r* by substituting estimates of the covariances and variances based on a sample into the formula above. That formula for *r* is:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$