UNIVERSITY OF
CAMBRIDGE

The Relevance of Real-time in Urban Digital Infrastructure

Technology and Data in Future Cities Workshop
22nd June 2018

Ian Lewis
ijl20@cam.ac.uk
Department of Computer Science and Technology
University of Cambridge

Questions:

How relevant is time in urban sensor data?
Is real-time relevant?
If so what about accuracy and latency?
Does moving from dozens to thousands of sensors make any difference?
Can we produce *generally* useful sensor analysis given the general characteristics of sensor data?

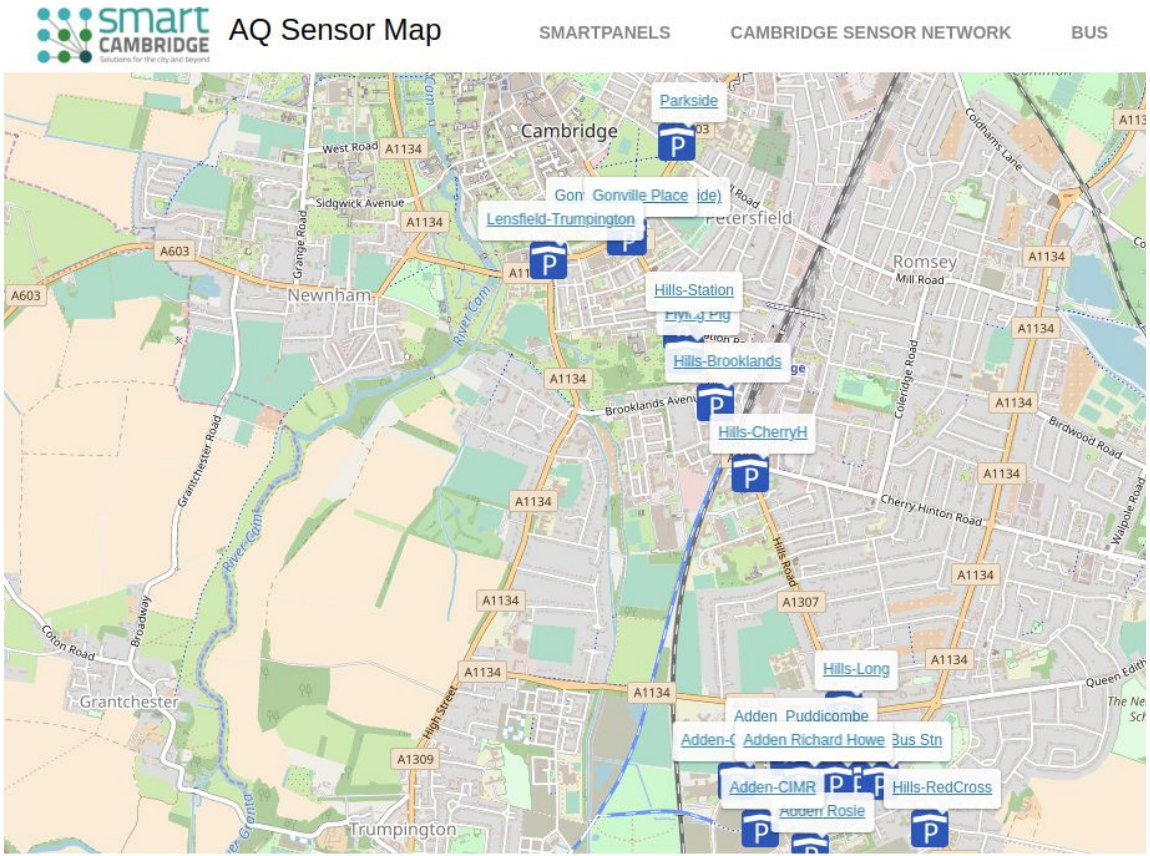
Answers:

No prizes for:

- Google
- Machine Learning

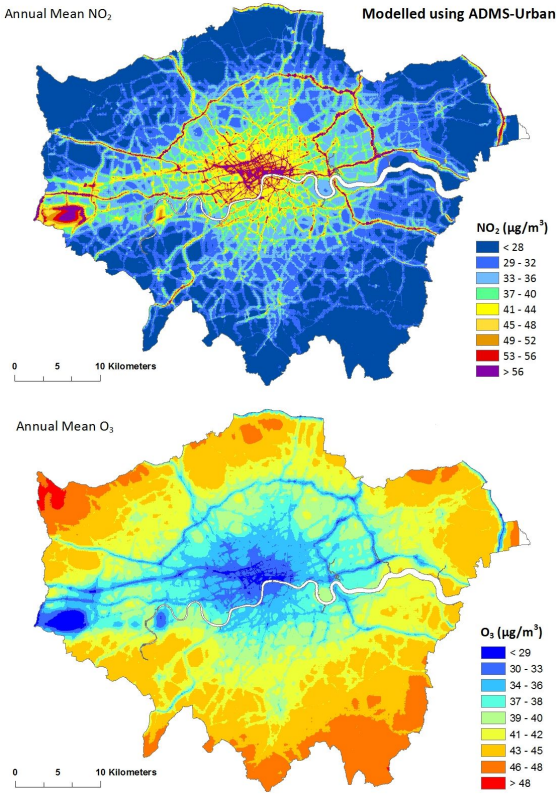
Current state-of-the-art: Sensor installation x dozens

Sensor installation x dozens - Air Quality in Cambridge 2016/2017



Rod Jones
Chemistry
University of Cambridge
2017

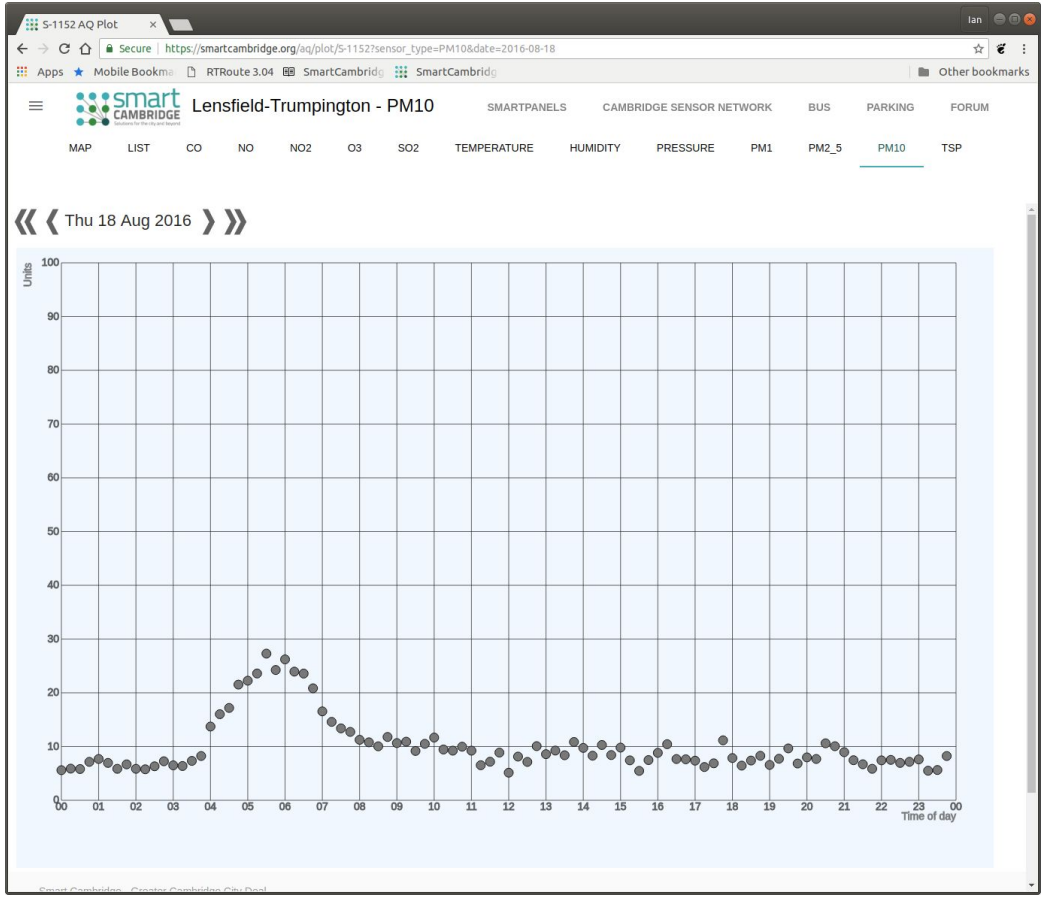
This data is (currently) spatially sparse, but analysis such as below is typical.
London NO2, O3 - CERC:



Contour plot of London showing the annual average NO2 and O3 concentrations predicted by ADMS-Urban for 2008. NO2 regions shown in yellow, orange or red are predicted to exceed the UK NAQS targets.

http://www.cerc.co.uk/environmental-software/ADMS-Urban-model.html

Air Quality: Data is temporally dense for a given location



PM-10 µg/m^3

Cambridge

Lensfield Rd / Trumpington Road

18 Aug 2016

Rod Jones
Chemistry
University of Cambridge

Kolkata Metropolitan area PM-10 µg/m^3

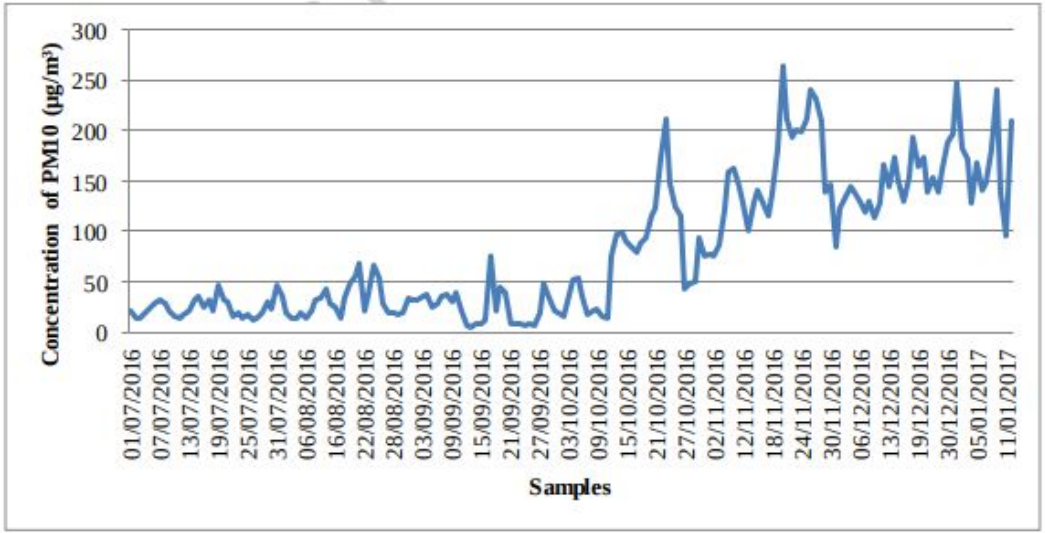
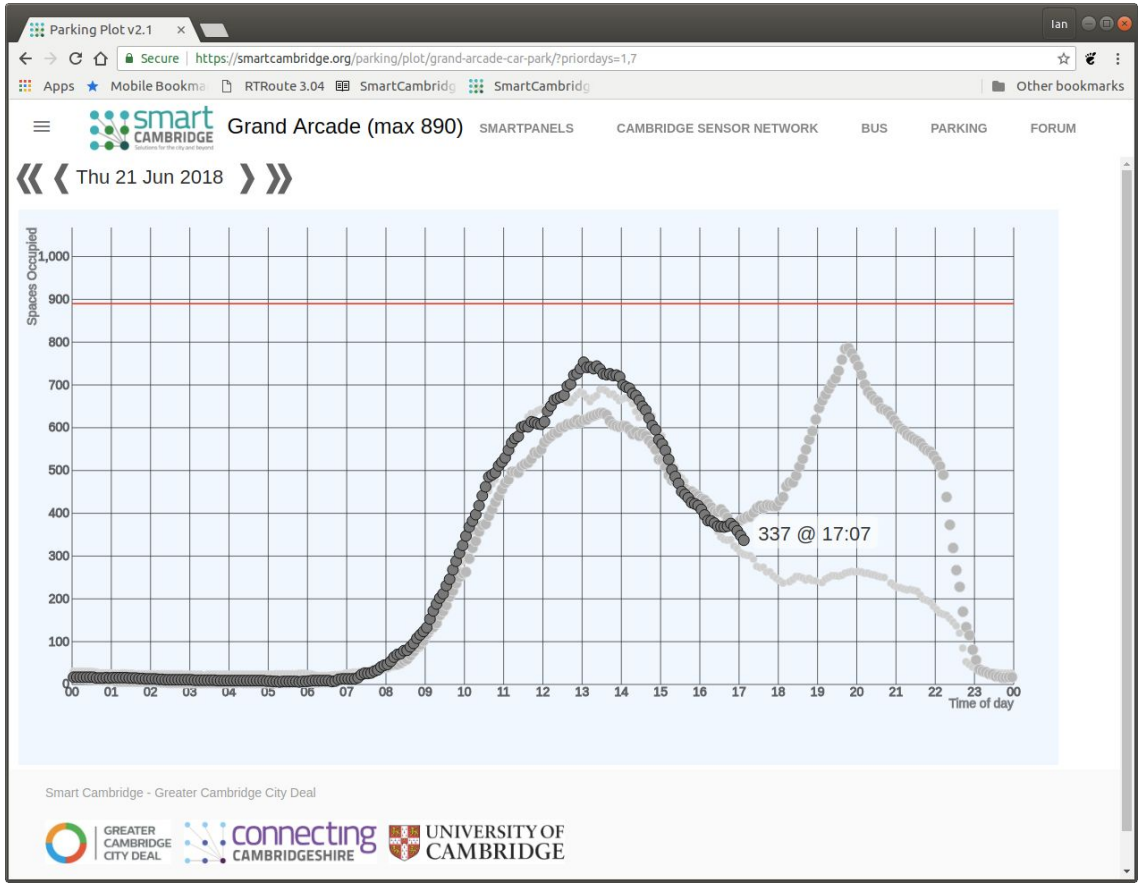


Fig. 5. Daily variation in PM_{10} concentration.

Air quality assessment using weighted interval type-2 fuzzy inference system
Joy Debnath, Debasish Majumder, Animesh Biswas
Ecological Informatics 15 June 2018

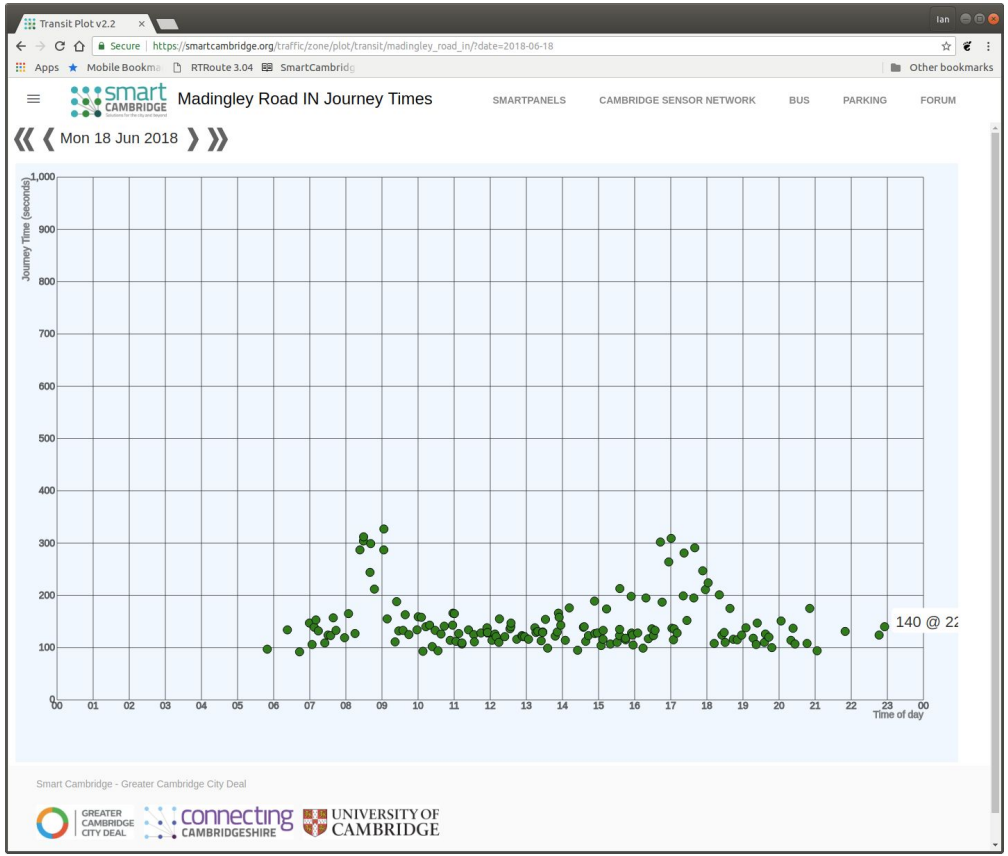
Car Parking: Data is temporally dense for a given fixed location



Space occupancy
Grand Arcade Car Park

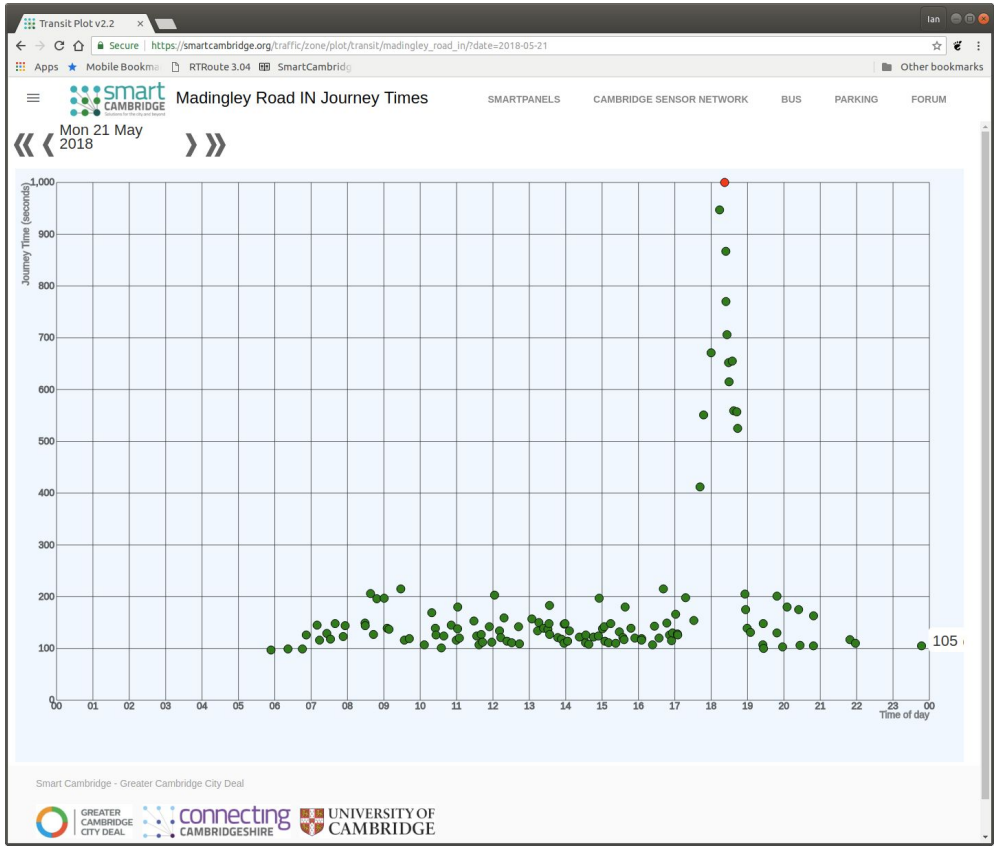
yesterday

Journey times: this might be a typical day on Madingley Road



Monday...

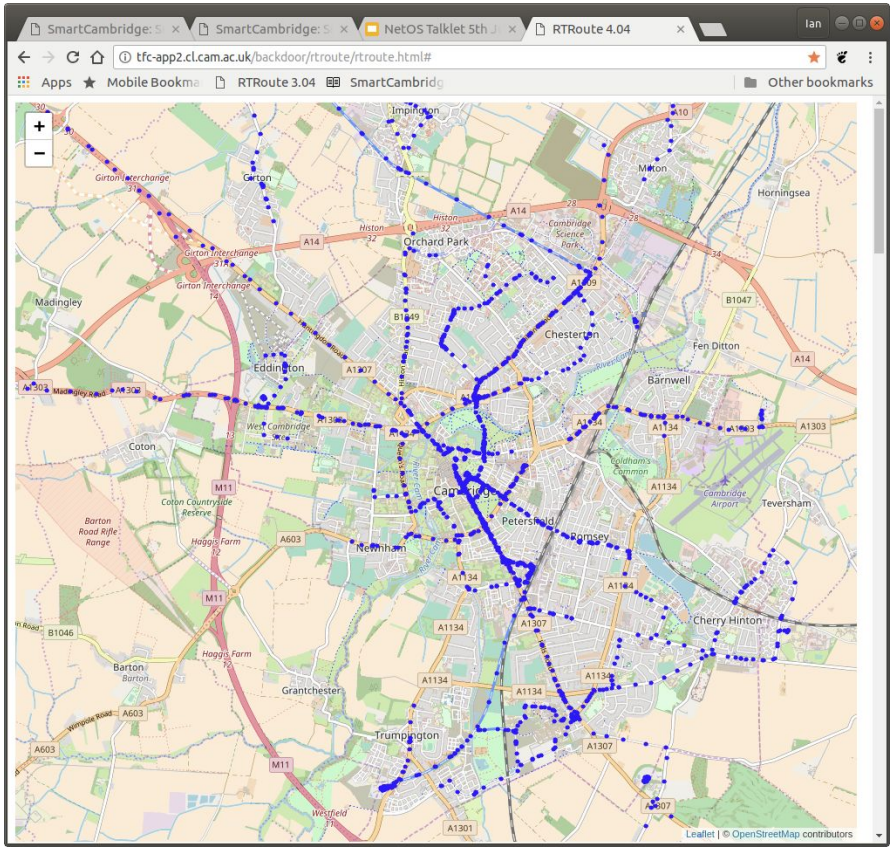
Journey times: but sometimes it looks like this



Monday…
A month ago

When might you decide to act upon data?

Moving sensors - spatially dense, temporally sparse



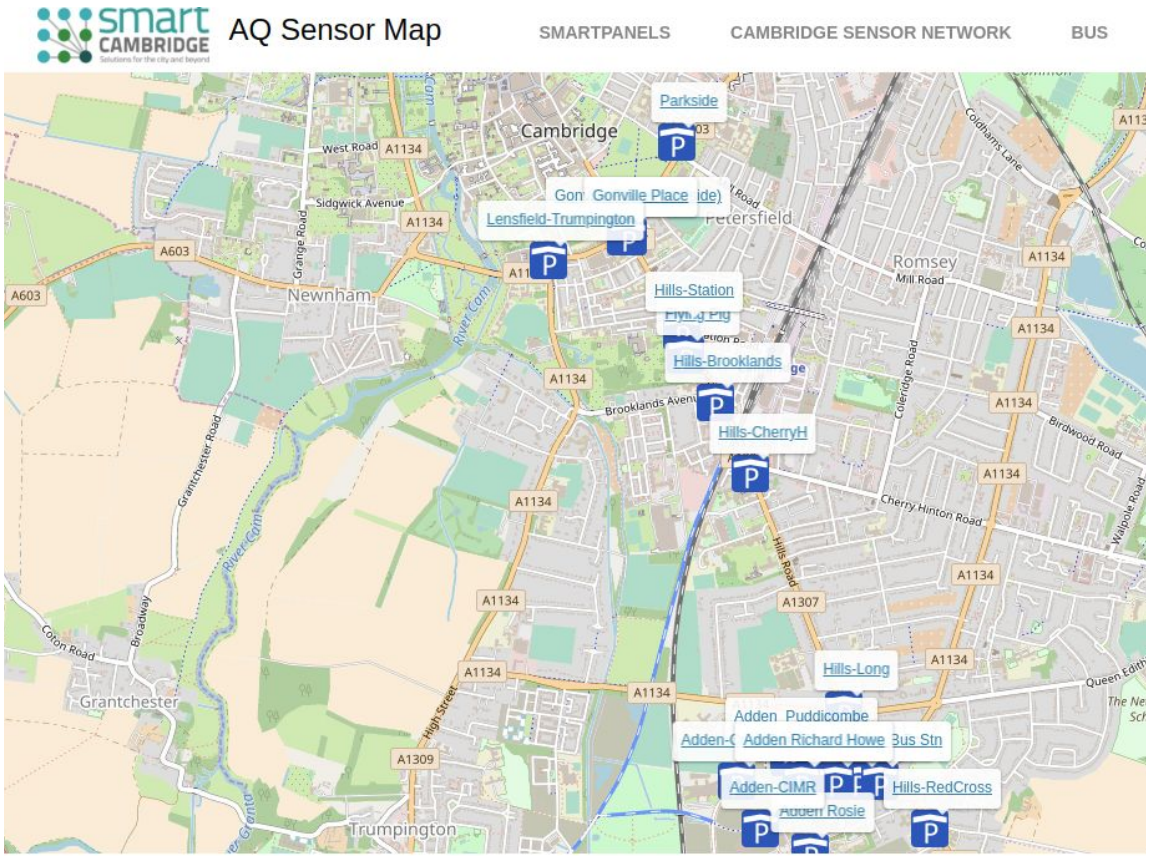
Cambridge

Bus movement data

~10 minutes

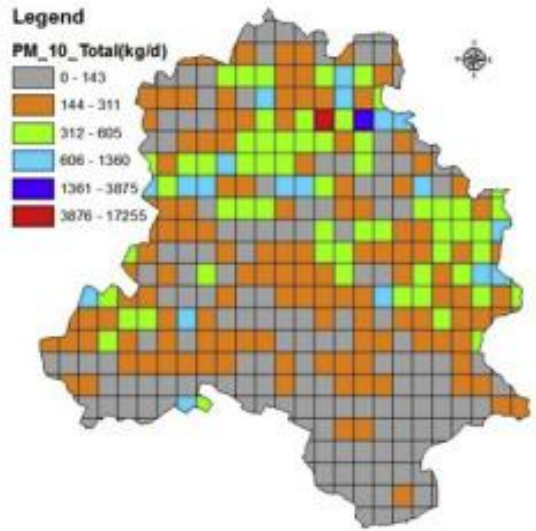
recently

Spatially sparse



Rod Jones
Chemistry
University of Cambridge
2017

PM-10 distribution in Delhi, 2010 (kg/d) - extrapolation from 10 sensors to 400.



Managing future air quality in megacities: Co-benefit assessment for Delhi

Anil D.Bhanarkar, Pallav Purohit, Peter Rafaj, Markus Amann, Imrich Bertok, Janusz Cofala, Padma S.Rao, B.Harsha Vardhan, Gregor Kiesewetter, Robert Sander, Wolfgang Schöpp, Dipanjali Majumdar, Anjali Srivastava, Swapnil Deshmukh, Amit Kawarti, Rakesh Kumar

Atmospheric Environment Volume 186, August 2018, Pages 158-177

So should you care about the time at which the data was recorded, or should you care whether the processing of the data should be near to real-time?

“It all depends what you want the data for”

We are proceeding on the assumption that the primary use case in an urban environment is we are measuring something because we want to do something about it. For example we want timely actions to occur to mitigate issues that are happening or predicted to happen. As levels of automation increase, this is better described as the city *ADAPTING* rather than an *INTERVENTION*.

Hence the *Adaptive Cities Programme* in the Cambridge Dept. of Computer Science.

The Adaptive City

Sensors

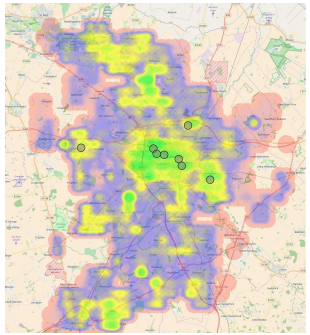
Sensor Network

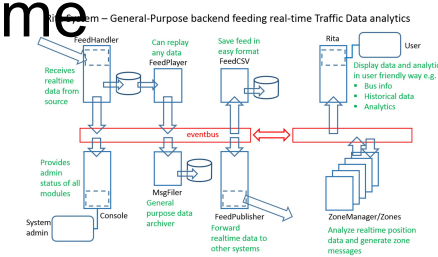
Real-Time

Storage

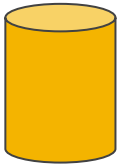
Analyse

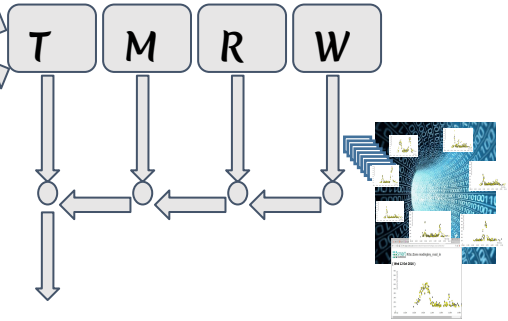
Predict





Adapt



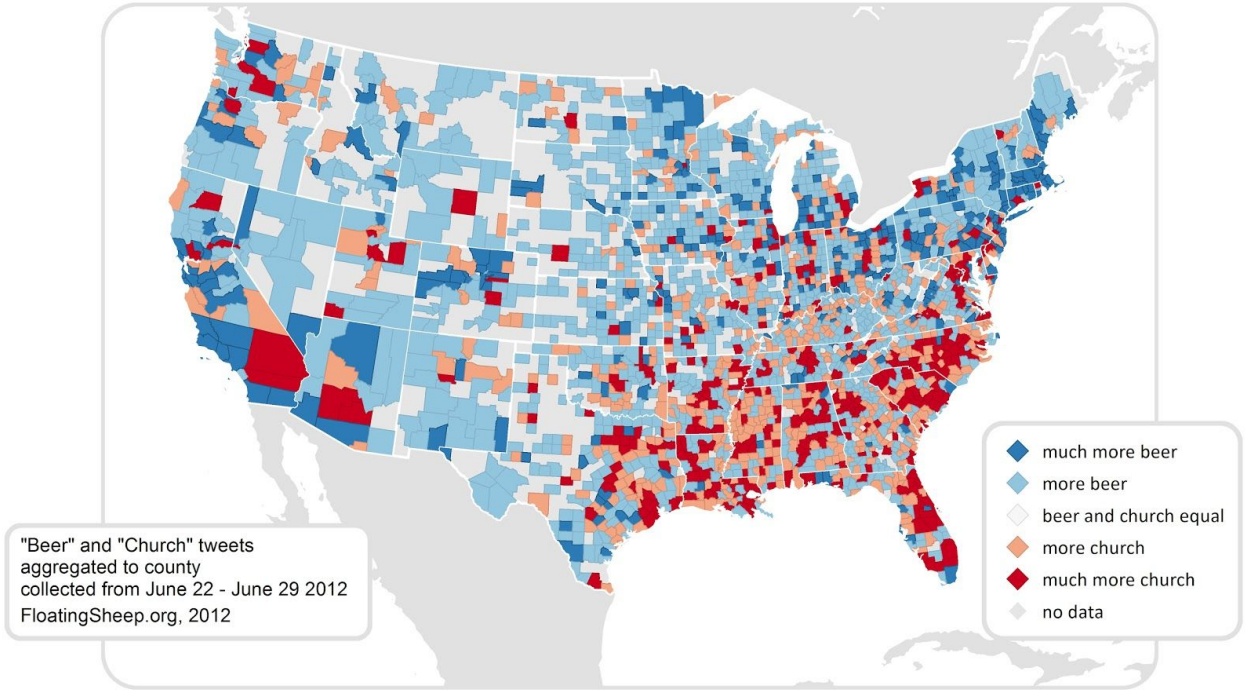


More about time...

All the data we are looking at is geolocated & timestamped:

(latitude, longitude, altitude, time, ...other stuff …)

Twitter as a data source, Church vs Beer, USA
FloatingSheep.org 2012

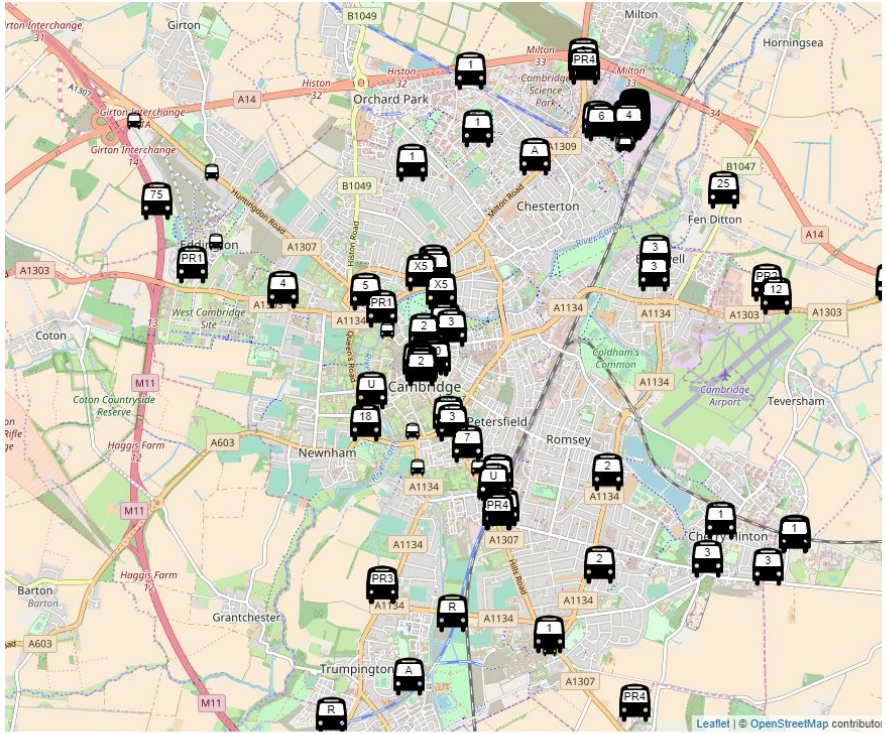


(latitude, longitude, altitude, time, ...other stuff …)

Format data as GeoJson, import into GIS package, draw previous map

But GeoJson has no concept of time.

Ian’s favourite sensor, the Bus:



Cambridge, 7pm yesterday…

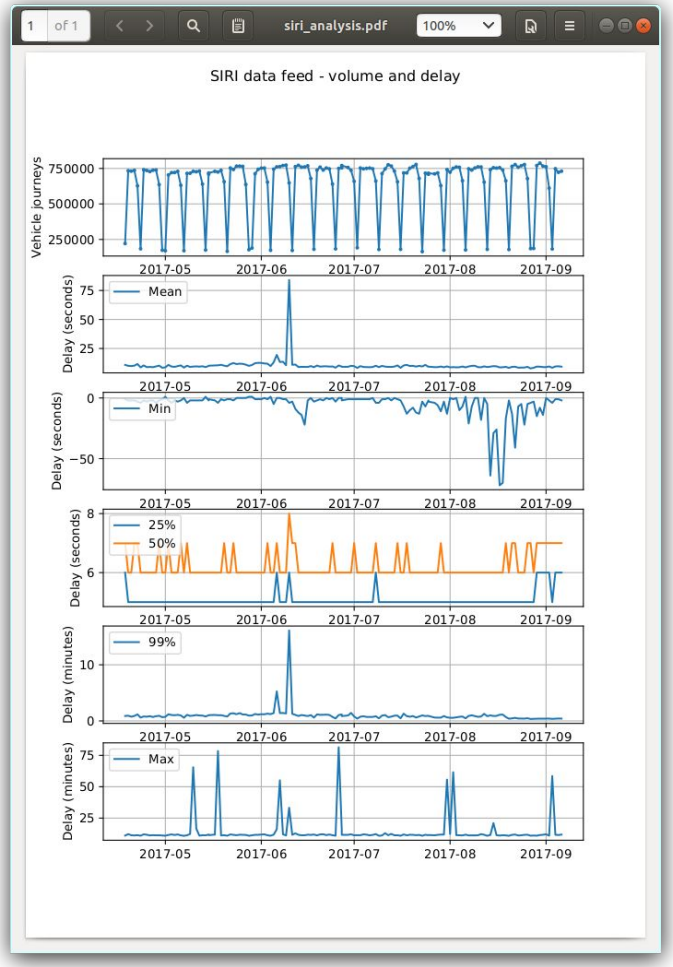
There is a great deal of useful analysis you could do with bus position data...

(latitude, longitude, altitude, time, ...other stuff …)

Talking about time…
(Buses have a timetable…)

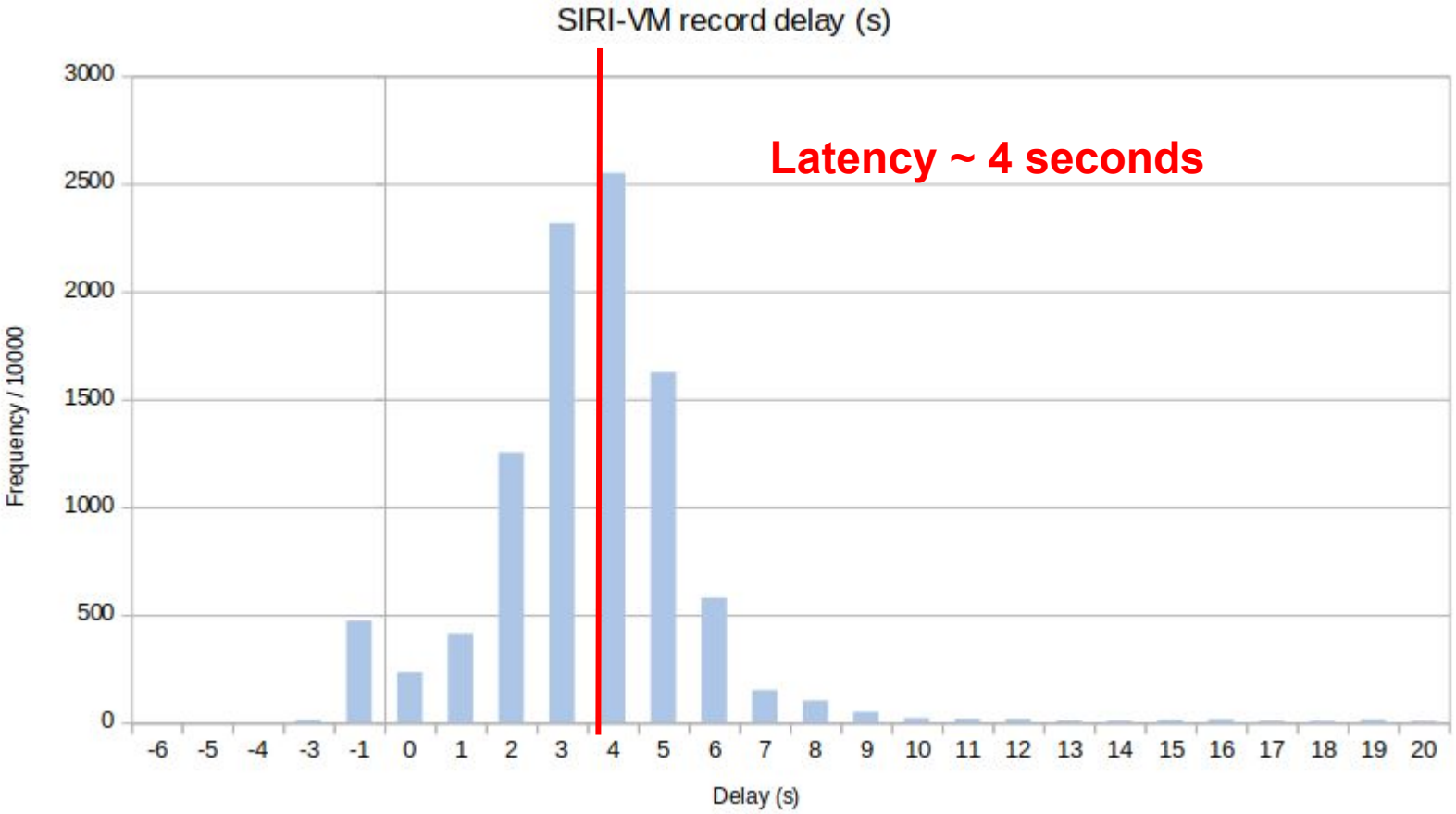
Cambridge region bus journeys
May-Sept 2017
~ 100 million data records

Orange line: median ‘latency’ ~ 6 secs.



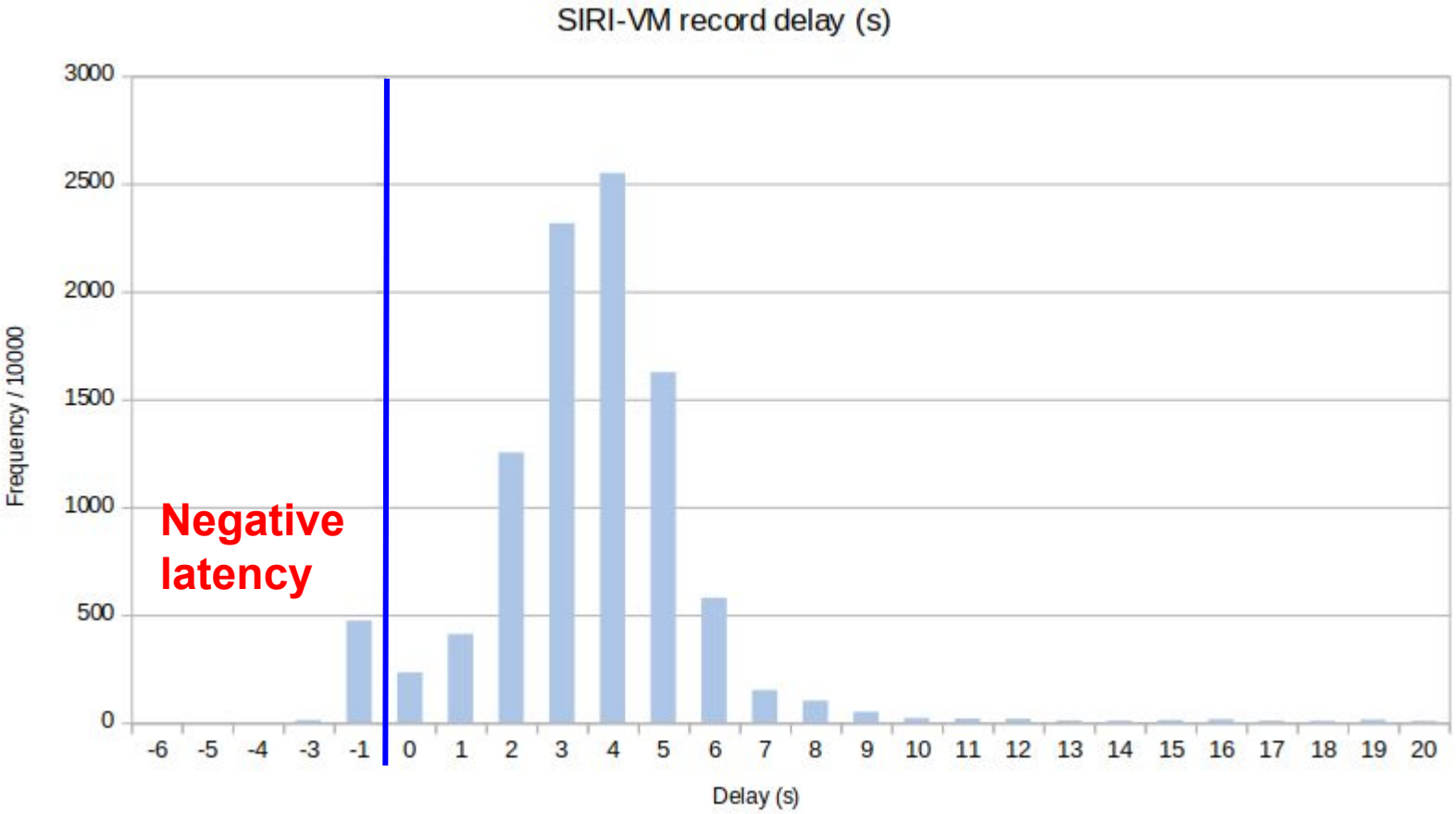
J Warbrick
Cambridge
Sept 2017

More recently, small snapshot (10,000 records) during an hour.

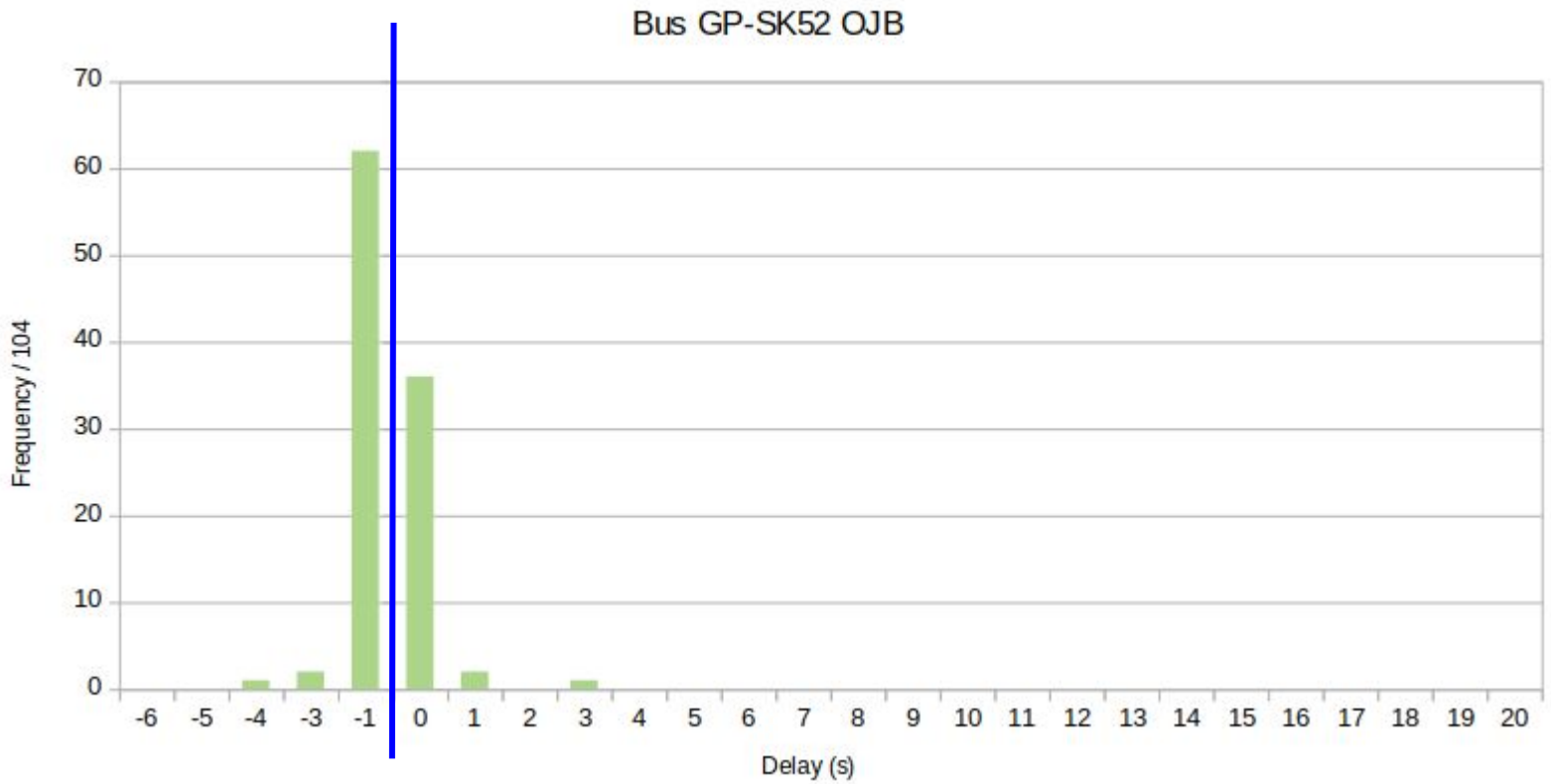


An example single bus

Back to 10,000 records, note sub-zero latencies

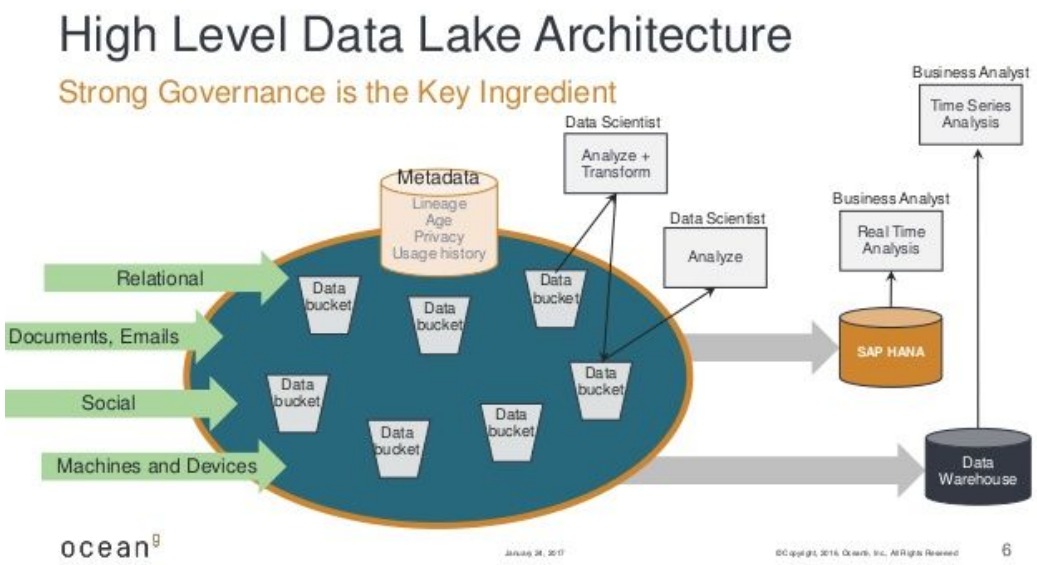


Need careful consideration of latency and accuracy



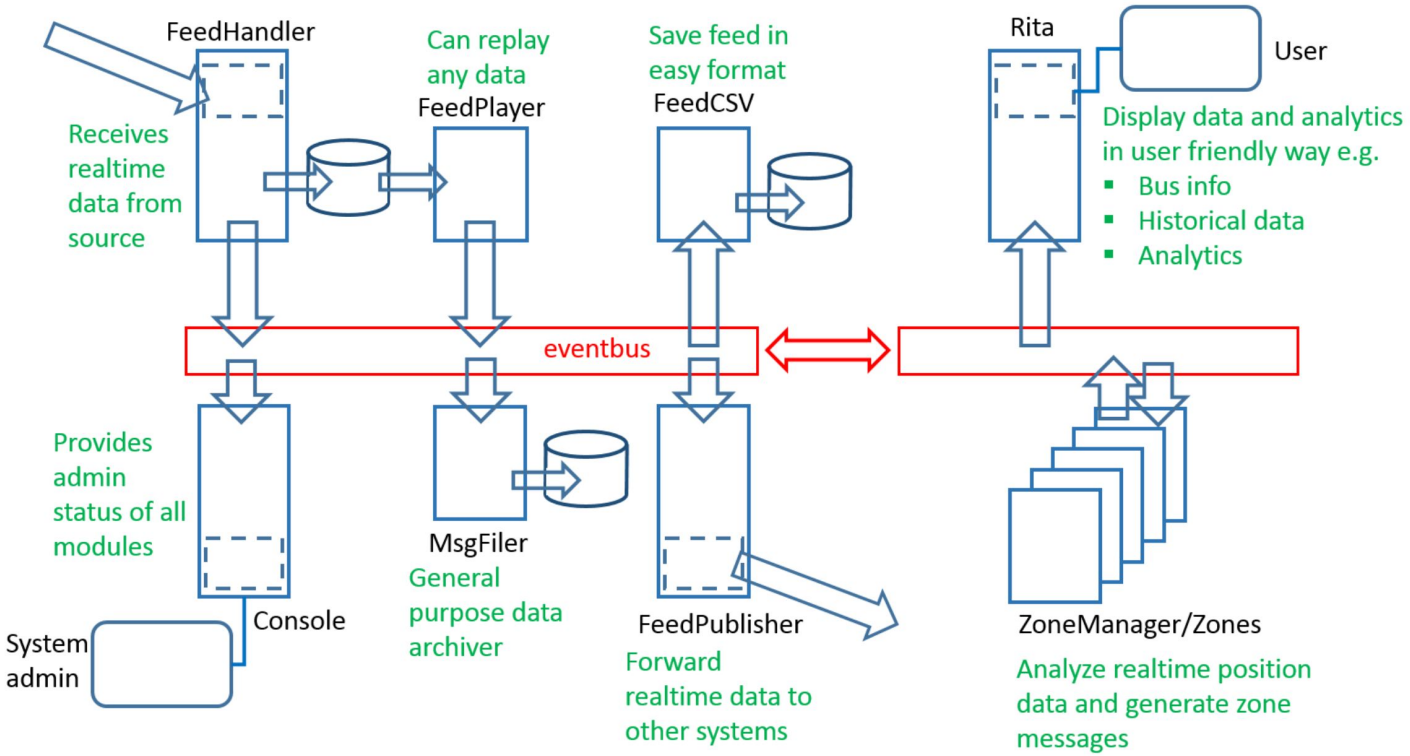
OK, so you’ve installed your sensors:

Does Real-time Really Matter? Smart City = Big Data

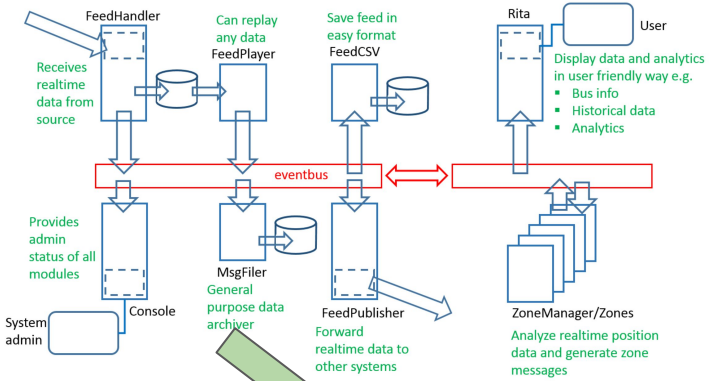


ref. Ocean9, Inc.
“With the growing implementation of IoT solutions resulting in ongoing streams of data, the data lake concept provides the right amount of rigor, without limiting flexibility as well as broad and open usability in a nevertheless governed fashion.”

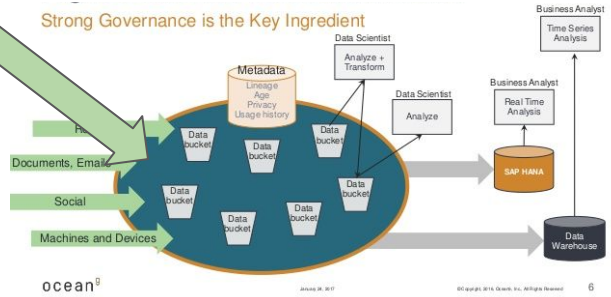
Cambridge Intelligent City Platform



Plausible to add storage to a real-time platform, Difficult the other way round.

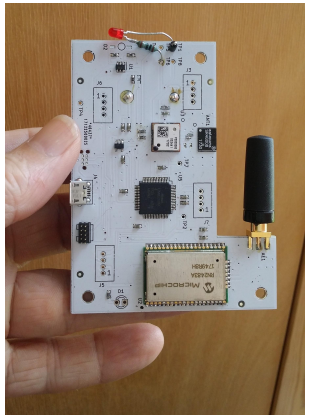


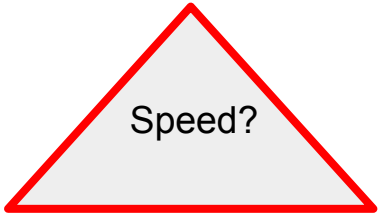
Do both ????



Time in the sensor network

1000 sensor deployment? LPWAN for the WIN



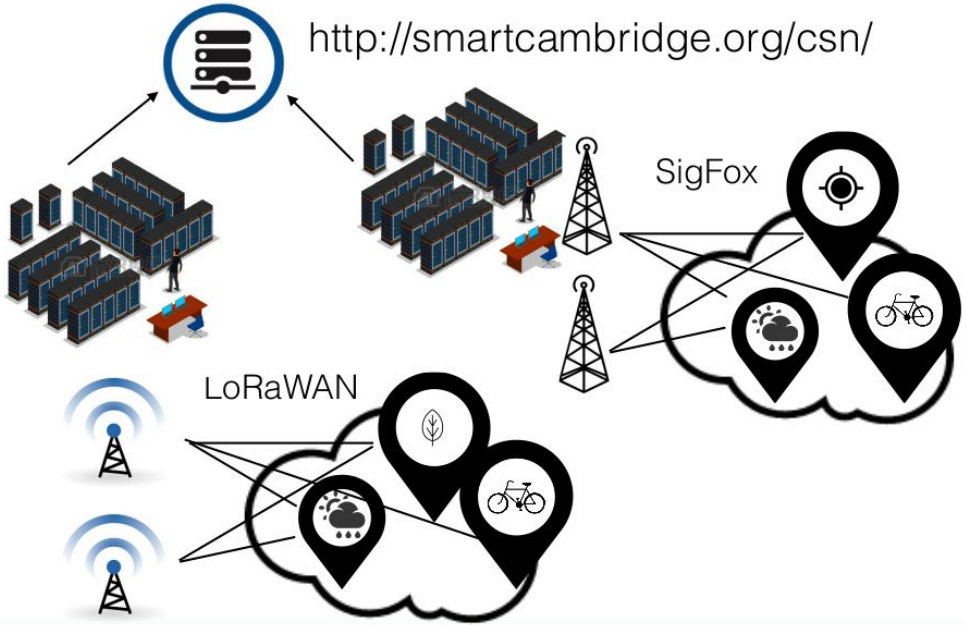


LoraWAN - public access (~ wifi), The Things Network. Radio made by Semtech (~$0.5)

SigFox - a single global network operator

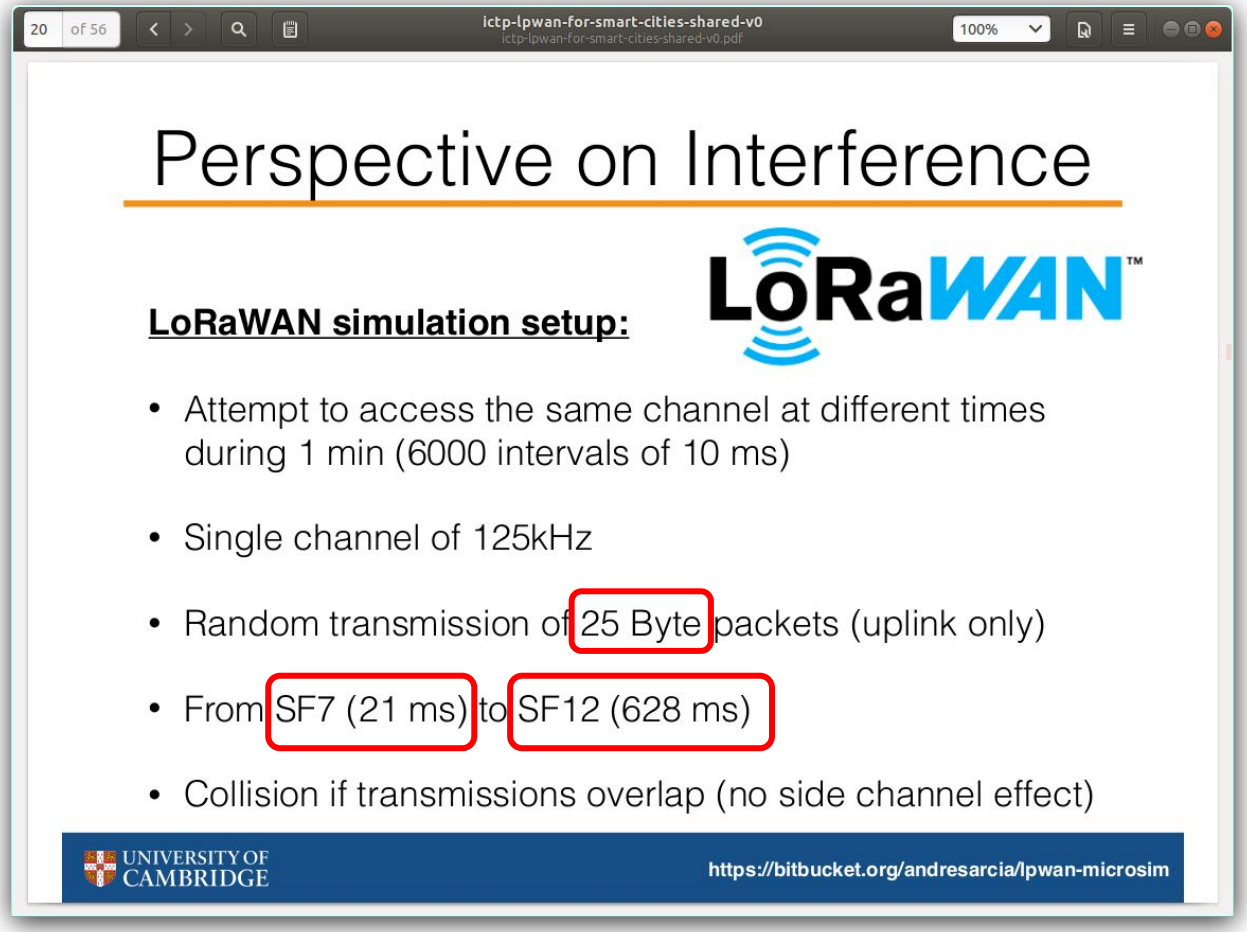
NB-IOT - the mobile operator solution

IoT Network Infrastructure



Andrés Arcia-Moret

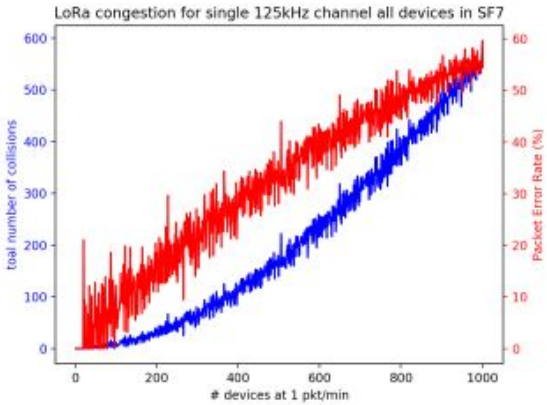
Cambridge
Dept. Computer Sci.
June 2018



Andrés Arcia-Moret

Cambridge
Dept. Computer Sci.
June 2018

Latency increases rapidly in a congested network

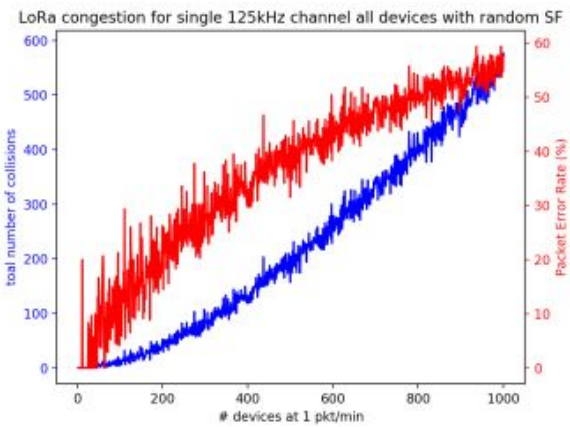


SF7

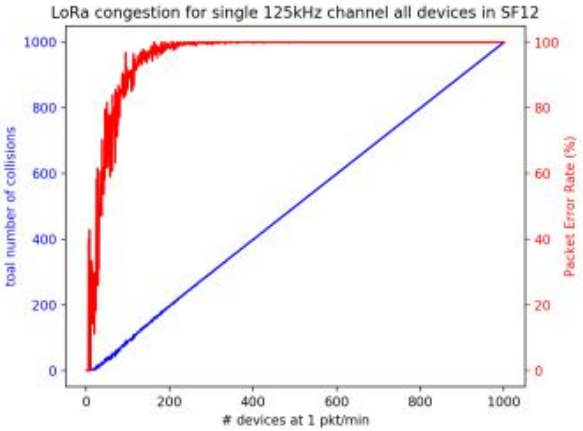
21ms

@ 1pkt/min

~100’s devices



random SF

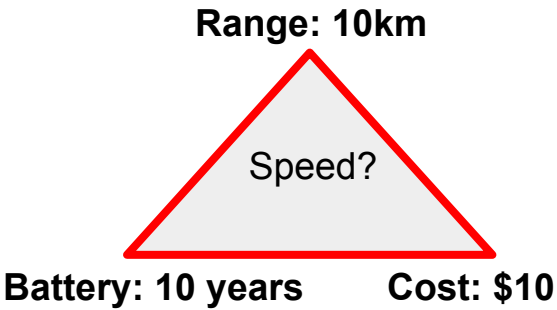


SF12

628ms

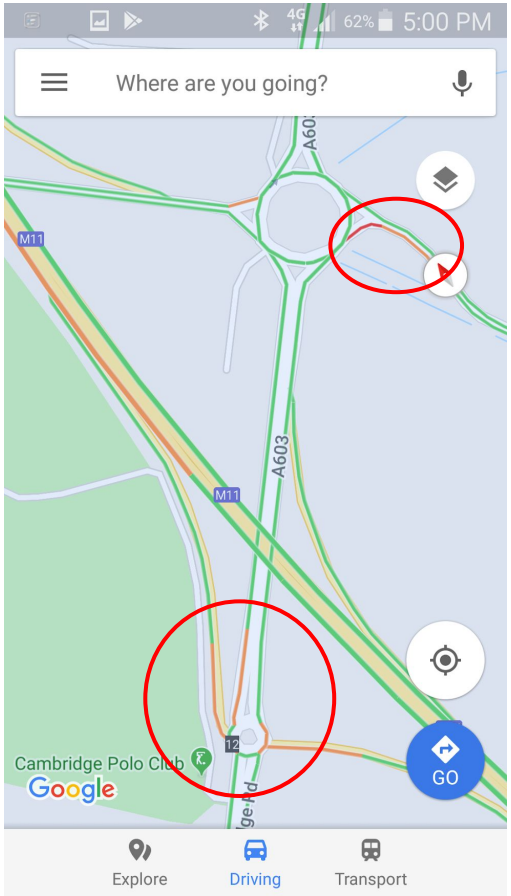
@ 1pkt/min

~10’s devices



Note that you need a network like this to deploy 1000’s of sensors...

Google - what's the latency, accuracy ?



Summary

The value of time in sensor data is not well recognised.
Current ‘Smart City’ initiatives place emphasis on collection, not processing.
Time becomes interesting when you correlate data.
Accuracy becomes important when you care about the time.
Accuracy, Latency become critical when you want timely action.
There are more ways to obtain pathologically wrong time data than you realise.
Consistent treatment as 4D data has merit.

Ian Lewis
ijl20@cam.ac.uk
Department of Computer Science and Technology
University of Cambridge