Big Data to improve the quality of operations: application to Crossrail in London
A new railway service across London from Est to West, going also outside the Greater London area.
A new underground link between existing rail lines through the construction of a new tunnel in the city center.

In collaboration

Surface existing line
Tunnel
- Existing Station
- New Station
- New infrastructure

Reading
Shenfield

Surface existing line
Tunnel
- Existing Station
- New Station
- New infrastructure

2018 RAIL CONFERENCE - Big Data to improve the quality of operations: application to Crossrail in London
Nine new stations and new connections with the airports, Network Rail stations, the Underground and Dockland LR.
A very dense service, with 24 trains per hours and direction during the peaks through the central section.

**Notes**

**Taplow and Burnham stations**
- C5 services only stop here in peak time.

**Liverpool Street station**
- C7 services depart from the overground main line terminus.

**Key to symbols**

- Change for Underground, Overground, DLR
- Change for National Rail
- Change for Heathrow Express
- For this station, see notes
- Services only operate at peak time

**Services**

<table>
<thead>
<tr>
<th>Services</th>
<th>Timetable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbey Wood – Heathrow</td>
<td>4 – 4</td>
</tr>
<tr>
<td>Abbey Wood – Paddington</td>
<td>6 – 4</td>
</tr>
<tr>
<td>Shenfield – Paddington</td>
<td>8 – 4</td>
</tr>
<tr>
<td>Shenfield – Reading</td>
<td>2 – 2</td>
</tr>
<tr>
<td>Shenfield – Maidenhead</td>
<td>2 – 2</td>
</tr>
<tr>
<td>Abbey Wood – West Drayton</td>
<td>2 – –</td>
</tr>
<tr>
<td>Gidea Park – Liverpool St</td>
<td>4 – –</td>
</tr>
</tbody>
</table>

**Occasional peak only services:**
- C13 Shenfield – Heathrow
- C36 Shenfield – West Drayton
- C24 Abbey Wood – Reading
- C25 Abbey Wood – Maidenhead

Crosrail services depicted as played by Transport for London line numbers are made up though.

This is not an official map. It was not authorised and is not affiliated with Transport for London National Rail or any other transport company or organisation.
3 signalling systems to be manage, a sequential opening in 4 stages.
How to ensure the best results and perfect operations for such a big and complex project?

MTR – Metro Hong Kong decided to use a dynamic simulation from the early tendering stage. The goal was to reduce the risks for the company ensuring a better quality of service during future operations. In this way it was possible to estimate (thus reducing possible malus and ensuring a better financial balance).

After winning the concession of Crossrail, MTR decided to continue to realize dynamic simulations in order to:

- Improve the quality of service on the first routes (on the existing tracks)
- Better prepare the future opening of the central section.
The rail planning can be viewed as a loop (see also alius presentation for the Rail Conference 2017), which starts from the problem definition and whose elements are the analysis of real data, the design of scenarios, the simulation and the analysis of simulated results. The loop is the ideal support to timetable planning, which is becoming more demanding since punctuality has become fundamental since delays lead to higher costs and lower customer satisfaction.
The approach for the simulation with the “planning loop” and the simulation model for Crossrail
The approach for the simulation with the “planning loop” and the simulation model for Crossrail
Goal for MTR in 2017:

- increase the quality of operations on the East part of the network
- reduce the risks for December 2018 (propagation of delays from the classic national network in the Crossrail tunnel)
How to use Big Data to improve the operations?

Workflow: main concept

Combination of stochastic simulation, analysis of train data, analysis of pax counts and smart card data and of course … experience ;-)

- Improve the timetable: Where?
- Expected Passengers
- Delays

New timetable

Evaluate the timetable

Improved timetable
How to use Big Data to improve the operations?

Workflow: input

Combination of stochastic simulation, analysis of train data, analysis of pax counts and smart card data, experience
How to use Big Data to improve the operations?
Utilization of the information about the passengers count

Just 2 interchange stations,
Interchange at Liverpool Street requires to exit the barriers
=> “Fixed” Routing
Passengers take the first train,
Some wait for the next one if the first too full

Aggregate the Demand
Demand/minute (Pax -> London/Station/min)

AFC data
How to use Big Data to improve the operations? Utilization of the information about the passengers count.

Oyster Data: Various Days

Oyster - Harold Wood

How to use Big Data to improve the operations?
Utilization of the information about the passengers count.
How to use Big Data to improve the operations?
Utilization of the information about the passengers count

- Stop time & N of Pax
- Relationship Pax Join/Al.
- Pax On Board - Dwell time

Graphs showing the relationship between stop time and number of passengers for different locations:
- Brentwood
- Harold Wood
- Romford
- Chadwell Heath

Equations and correlation coefficients for each location:
- Brentwood: $y = 0.1726x + 27.62$, $R^2 = 0.279$
- Harold Wood: $y = 0.0623x + 30.892$, $R^2 = 0.09344$
- Romford: $y = 0.0639x + 33.702$, $R^2 = 0.22196$
- Chadwell Heath: $y = 0.1113x + 32.935$, $R^2 = 0.55365$
How to use Big Data to improve the operations?

The relationships obtained for each station and displayed at the previous slide are used for the simulation. The line obtained is combined with a log-normal distribution representing the deviation from this deterministic mean.
How to use Big Data to improve the operations?

Workflow

- Relationship: Pax Join/Al.-Pax On Board - Dwell time
- Expected (Mean): Dwell time/Pax
- Delays

- Demand/minute (Pax -> London/Station/min)

- Critical Trains (Pax on board)

- Critical Trains Delay

- Timetable Draft

- Pax < Capacity?
- Robust? (PPM?)

- Improved Timetable
How to use Big Data to improve the operations?

Analysis of real operation data (berth level data)

Comparison between planned timetable and real occupation

![Graph showing comparison between planned and real operation data for different locations.](image-url)
How to use Big Data to improve the operations?
Analysis of real operation data (berth level data)

Punctuality (1’, 3’ and 5’ minutes delays) in a station during the morning peak hours
How to use Big Data to improve the operations? Identification of the most critical trains in the timetable.
How to use Big Data to improve the operations? Identification of the most critical trains in the timetable

Analysis of the actual data between **06:30 and 07:30**: late departure from GIDEA Park, then delays not recovered
Analysis of the actual data between **08:00 and 09:00**: Area of nearly 50% of the trains overlaps with the following one.
How to use Big Data to improve the operations?
Identification of the most critical trains in the timetable

Analysis of the actual data between 08:00 and 09:00:
Detail of a very critical train departing from GIDEA Park

The trains start to late and the running/stop time is not sufficient.
How to use Big Data to improve the operations?

Proposed timetable improvements

December 2015
Proposal
How to use Big Data to improve the operations? Estimation of the results with the simulation
How to use Big Data to improve the operations?
Estimation of the results with the simulation

More balanced use of trains
No train >1200 pax
How to use Big Data to improve the operations? Estimation of the results with the simulation

--- December 2015

Proposal

HPM: Proposal vs Dec 2015

How to use Big Data to improve the operations?

Estimation of the results with the simulation

December 2015

Proposal
How to use Big Data to improve the operations?  
Ex-post verification: April 2017

Comparison between planned timetable and real operation

![Comparison chart between planned timetable and real operation](image-url)
How to use Big Data to improve the operations?

Ex-post verification: May 2017

Comparison between planned timetable and real operation
How to use Big Data to improve the operations?

Ex-post verification: April 2017

Evolution of the performance along the line: train arriving at 08:35
How to use Big Data to improve the operations?
Ex-post verification: May 2017

Evolution of the performance along the line: train arriving at 08:35
How to use Big Data to improve the operations?

- Efficient and effective utilization on a synchronic dynamic simulation
- A good calibration of the model based on real data is essential to ensure good results
- The utilization of passenger data permit to ensure a better evaluation of the new dwell time at the stations. This value is very important in a very dense and saturated network.
- The utilization of a powerful new “planning suite” reduce significantly the project time and increase the accuracy of the results.

Conclusion

- The utilization of passenger data permit to ensure a better evaluation of the new dwell time at the stations. This value is very important in a very dense and saturated network.
- The utilization of a powerful new “planning suite” reduce significantly the project time and increase the accuracy of the results.