A scalable agent based multi-modal modelling framework using real-time big-data sources for cities

Gerard Casey, Corresponding Author
University of Cambridge
Department of Engineering, University of Cambridge, UK
Phone: + 00 44 77 61 31 97 49
Email: gac55@cam.ac.uk

Kenichi Soga
University of California, Berkeley
Department of Civil and Environmental Engineering, University of California, Berkeley, USA
Email: soga@berkeley.edu

Elisabete Silva
University of Cambridge
Department of Land Economy, University of Cambridge, UK
Email: es424@cam.ac.uk

Peter Guthrie
University of Cambridge
Department of Engineering, University of Cambridge, UK
Email: pmg31@cam.ac.uk

Krishna Kumar
University of Cambridge
Department of Engineering, University of Cambridge, UK
Email: kks32@cam.ac.uk

Word count: 5,998 words in abstract, text, references + 5 figures x 250 words (1,250) = 7,498 words

Submission Date
1st August 2016

ABSTRACT

This paper presents a framework for using real-time big-data to inform a transport Agent Based Model (ABM) for a range of scenario testing applications. Computational advances have enabled for increasingly complex, bottom-up, fine resolution simulations to be carried out over long time horizons at fine spatial and temporal resolution. This has hinted at the possibility of connecting scales of what has been historically been fine resolution operational models and coarse resolution strategic models. The value of any fine resolution dynamic model is limited by the quality of its inputs. The wave of new geospatially connected devices has enabled the harvesting of fine resolution spatial and temporal data on travellers' and even the infrastructure itself. This crowd-sourced data can be used to inform dynamic models with real-world and real-time data, bypassing the need for generalised functions and/or expensive survey data. In this paper, Google Directions API data and Transport for London data feeds are presented in a framework for London. The use of decentralised data structures is also presented and comment is made on the possibilities of using parallel computing advances in Computer Science to scaling up fine resolution scenario testing transportation models and enabling support for a range of agent decision making methodologies. Such data structures offer performance improvements in the storing of dynamic data that may be manipulated in order to simulate local and global hard infrastructure scenarios alone or in tandem with traditional policy or dynamic policy making scenarios.

Keywords: Big-data, real-time, traffic, GPS, modelling, HPC

INTRODUCTION

Historically, transport planners have made use of top-down macro-economic models that treat transport modes as distinct. These models were informed with static averaged inputs from standardised functions or limited survey data.

In recent years two profound changes have occurred. First, computational advances have enabled for complex, bottom-up, fine resolution simulations to be carried out over long time frames. Although the theoretical foundation for such dynamic models has existed for many years (*2, 3*, *26*,*34*) it has not been possible to implement them in a useful fashion until relatively recently. Secondly, the same computational revolution has created a network of geospatially connected devices which has enabled the harvesting of fine resolution spatial and temporal data on travellers' and even the infrastructure itself. This crowd-sourced data can be used to inform dynamic models with real-world and real-time data, bypassing the need for generalised functions and/or expensive survey data.

This paper proposes a framework which consists of a multi-modal agent based model which can be used for a range of scenario testing exercises. These may be hard infrastructure changes or policy changes. The agent based model makes use of a large repository of spatially and temporally dynamic data in order to provide real-world inputs.

An overview of the current models and approaches is first given. This discusses the backdrop to the emergence of bottom up dynamic models and the advent of crowd-sourced big data model inputs. If and how a traveller uses such information in their decision making follows. The use of a complex model with a large repository of input data can lead to large computational demand and thus scaling of this framework is discussed and finally a summary of the framework presented in this manuscript is given.

OVERVIEW OF MODELS & APPROACHES

Agent Based Modelling

Agent based modelling (ABM) has emerged as a means of dynamically simulating complex systems in a bottom up, stochastic method, rather than a deterministic top down method as has traditionally been advocated. The basic principle of an ABM is that discrete agents with distinct behaviours interact to bring out macro behaviour (*25,26*). A dynamic simulation can allow for the system processes analysed at the level of their constituent elements (*7*) and thus can permit a better understanding of the agents involved, their stochastic and heterogeneous attributes, and how their complex interactions lead to exhibited macro level behaviour (*14*). Recent advances in computational capacity have enabled more complex, dynamic simulations to be possible (*1*).

Model Inputs

In order to reflect real-world conditions accurate model inputs must be provided. Transport networks are complex and offer a multitude of options, via many different transport modes, for travelling to and from any location. The network varies considerably relative to spatial location, with some areas being well connected to the rest of the network and other areas considerably less so. Such connectivity may be defined as an output from the minimum cost route between the origin and destination. Generally, this minimum cost is defined as a combination of the monetary and non-monetary costs of this journey. A given traveller may make their travel decision based upon a range of different network related metrics and attribute weights to those metrics in a simplistic or complex fashion. For example, the value attribution of different modal time weights may differ, with waiting times discounted more than in transit times (*13*). Such discounts may be dependent on the socio-economic status of the traveller, the time of day, the travellers' role or even the weather at the time of travel (*4*). This section begins by explaining how metrics have historically been quantified for the transport network.

Road Network

Harvesting road vehicle related data for all roads has historically been prohibitively expensive and often standardised functions of sample roads are used to find suitable values (*27, 20*). Geospatial data such as road length, lane count, road type and survey data such as traffic counts has enabled the use of generalised functions such as bimodal journey time functions (*19*) and volume-delay functions in order to estimate likely road attributes. Such functions are derived from limited, old and extremely context specific studies resulting in a limited empirical evidence base which is increasingly far removed from the modern context (*24,29*). Traffic counts are often converted to Annual Average Daily Flows (AADFs) carried out over short periods and averaged over long periods (*30*), offering a limited snapshot and little in the way of temporal distribution.

These functions attempt to generalise different aspects of a roads characteristics in order to create general functions without the need for input surveys. However, in doing so their ability to give outputs that consider the context specific nature of of a given road reduces. Such differing characteristics can result in very different vehicular behaviour on roads that may be considered similar by these functions (*6*).

Public Transport Network

Public transport timetables provide a centralised resource for quantifying the journey time and financial cost attributes of public transport services. Public transport services are centrally coordinated and scheduled in advance in contrast to the decentralised/individual nature of most car journeys. In the case of cities where a centralised body is responsible for public transport services it is often possible to access all public transport mode data through one centralised repository.

Simplifications are usually employed in order to consider how a traveller may be presented with a particular service, for example journey times often include half the head time between services to give a static journey time that considers scheduling (*9*). Such assumptions negate the identified impact of different timetables that is known to influence a traveller's view of a public transport service (*9*). It also fails to consider the reliability of services and their tendency to provide a level of service as is specified in the timetable. Many transport systems in major cities are stressed at times, often resulting in significant service impacts for travellers'. For illustration consider the service performance of the London Underground service (*http://tubestatus.net/graph*). The result may be highly variable service reliability that can have a resulting impact on traveller decision making. The difference between planned public transport services and actual public transport services may have an impact on the robustness of using the idealised timetable as a model input.

Summary

The ability of any model to accurately reflect a real-world decision is dependent on the inputs it is provided. In the case of road and public transport journeys it is well known than context specific road conditions and timetables lead to highly time dependent journey times that travellers, are to varying degrees, aware of (*9*).

Despite this, transport models have historically used static, one-point inputs, such as traffic counts for one day, on one discrete part of a large network (*8*). Therefore, such methods are unable to capture information at the granularity or at the correct scale to accurately quantify how the transport infrastructure performs over time, both locally and globally. These metrics and their associated variability/elasticity that control the behaviour of the system are also unknown. Building and maintaining information on hard infrastructure is extremely time and resource intensive and yet it would be a relatively simple task in comparison to building the same for temporal information such as that caused by traffic congestion, which can fluctuate minute by minute (*33*). Capturing such temporal variations were a key consideration of this research and a range of different data sources were investigated.

MODEL INPUTS

In the case of London there is a road network, with varying road types, a rail network, with Overground and Underground services, the tram network and even river boat network. Each of these individual networks can be considered as a graph with individual properties and behaviours. A graph is an abstraction of the network in reality in the form of linked vertices. It is possible to transfer between different graphs at defined points, for example a traveller may move from the road network to the rail network at a train station. The result is a multi-layered transport graph.

Model Inputs

Fundamentally, the framework supports three distinct types of geospatial data:

- Point data
- Line data
- Polygon data

Point (vertex, node) and line (edge, link) data collectively form a graph. This geospatial data is either temporally static (e.g. road length) or temporally dynamic (e.g. journey time). Figure 1 illustrates these data types and provides illustrative examples.

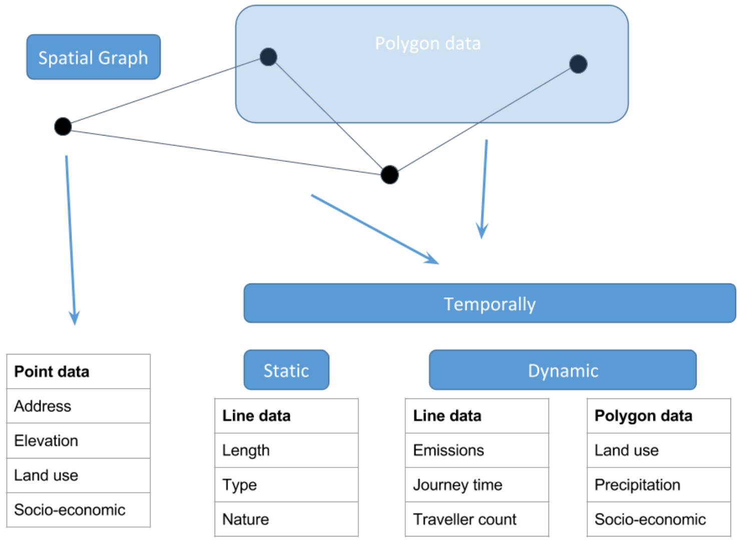

FIGURE 1 Data types

Graph Permissions

In order to manage the differing attributes of different services it is necessary to stipulate a hierarchy of graph types with common attributes. These attributes specify basic properties of each graph and ensure that only permitted agents are able to utilise different graphs.

Road Graph

The UK Ordnance Survey (OS) Integrated Transport Network (ITN) (*22*) was used as the base map for the road network. Postcode, street name, administrative areas *etc.* were taken from the Google Geocode API & OS ITN, elevation data from the Google Elevation API and socioeconomic metrics from census data from the Office of National Statistics (*21*) and land use from the ONS.

Public Transport Graph

A public transport network was constructed by combining separate rail (overground, tram, underground and National Rail) shapefiles into one connected graph. Interchange times between platforms within stations and other walking aspects of the public transport network are included.

Temporal data

This section covers the data which changes with time, for example link journey times.

Road Graph

Journey times on the road network in most major cities can be highly variable. GPS enabled mobile phones have enabled the harvesting of real-time empirical travel data, most especially in areas of high population density such as cities (*11,23*). Rather than using coefficient based generalised functions, real-time queries can be made over long periods of time, creating a historical database from which trends can be analysed. There are a range of providers who provide shortest path directions as a service, for example Apple, Bing, TomTom and Google. Generally, these services are targeted at users who wish to make a route choice or a modal choice for a given route or routes. The aim here was not to make use of the specialised shortest path algorithms or large-scale computational power of these service providers but rather to access GPS informed journey times.

This research made use of the driving side of the Google Directions Application Program Interface (API) (*15*) to harvest GPS informed journey times for all roads in the Greater London Area. The Google Directions API is a service that calculates directions between locations using a Hypertext Transfer Protocol (HTTP) request. This HTTP request can be formulated to poll at any given temporal resolution for any given origin and destination pairs, limited only by usage restrictions. In this case, an array of origin and destination pairs covering all of London were queried at 2 hour intervals (from 6am to 10pm) over a complete period of 2 months. Geographic output zone data was used to build this OD matrix.

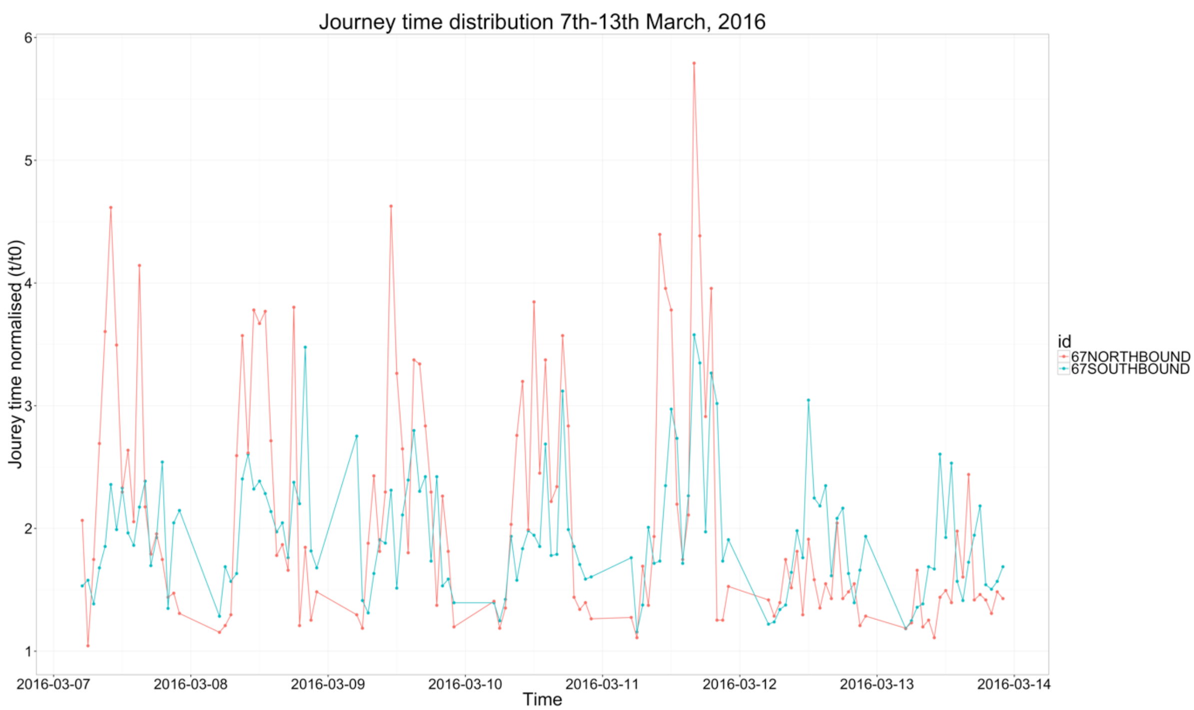

FIGURE 2 Journey time distributions over one week, Buckingham Palace Rd, London (*4*)

Figure 2 presents a graphical demonstration of the dynamic distribution of journey times on one sample road in London, in both directions over a period of one week. It is clear that the northbound (towards city centre) lane experiences the greatest increase in journey times in the mornings and the southbound lane experiences the opposite pattern of lower journey times in the mornings and longer journey times in the evenings. The weekly distribution also shows that no day is identical, with each day exhibiting a unique footprint. Friday shows a distinctive inverse of the other week days with the out bound lane demonstrating the highest journey times all week. Saturday and Sunday display a different trend to weekdays with the peak increases occurring in the outbound direction rather than the inbound direction which occurs on weekdays (*6*).

Due to the (legitimate) tendency for routes favoring main roads it is necessary to employ an iterative approach to achieve sufficient coverage of the graph with Google Directions API data. This requires the requests process to be informed of locations on the underlying graph that have either little or no temporal data in order for it to edit future requests to capture these locations.

After a period of iterative requests, it will likely be observed that some areas remain unpopulated. This is generally areas of either very low usage (thus no GPS data is available) or areas with poor reception (thus GPS signal is not available). In these cases, the travel time distributions from similar road types can be used to plug the gap. It is likely such areas of poor real-time data coverage will be of little strategic importance yet care must still be taken to assess the proportion of real-data to synthetic data in the underlying model. An iterative approach was deemed the most appropriate method for solving the coverage issue in order for this method to be used with other data providers and in other locations.

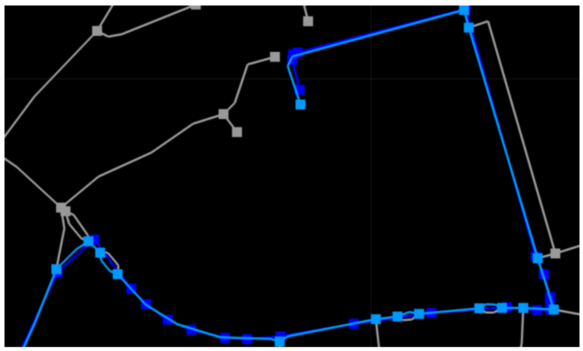

FIGURE 3 Resolving ITN Network with overlaid Google Directions API data

At this point, it is now possible to take journey times from the Google Directions API result and pass these journey times as attributes to the underlying ITN graph. Beyond graph coverage, there is also the issue of resolving the difference between Google's graph and that of the ITN. Consider Figure 3 which presents a portion of the ITN graph and a Google Directions API request polyline which intersects this area. The dark blue polyline illustrates the Google Directions API result and the grey illustrates the underlying ITN network. As is visually clear, there are small discrepancies between Google and ITN's underlying graphs. It is necessary to reconcile this two differing representations of the same physical infrastructure in order to transfer attributes from the Google data (journey times) to the ITN graph. A method using a shortest distance path implementation with weights relative to vertex proximity was employed. A quad tree data structure was also employed for computational efficiency. The result is the light blue polyline in Figure 3. Thus, as a batch process the attributes from the Google data can be matched to the underlying ITN graph and at this point the journey time attributes can be transferred from the Google result to the appropriate ITN edge.

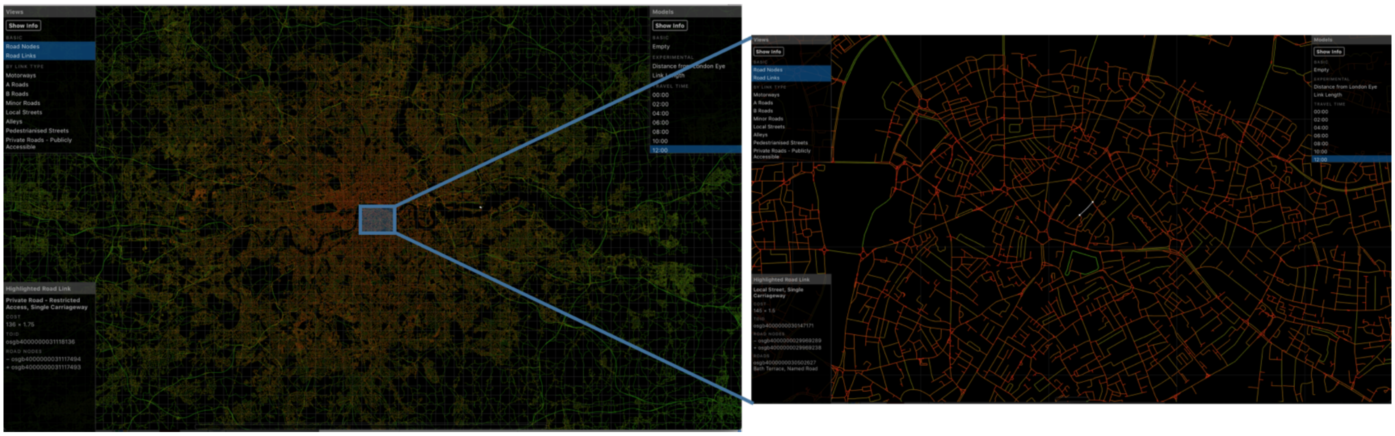

FIGURE 4 Visualisation of journey times in web browser interface at midday 10/23/2015. Red indicates an increase in journey times and green indicates a lower journey time relative to free flow journey time. The image on the left displays a macro view of London. The inset is a micro view of the area of Southbank and Newington.

The result is a road graph populated with temporally dynamic journey times as is visualised in Figure 4. Other attributes may be derived from these journey time attributes. For example, estimated financial costs, vehicle counts and emissions. Again, different methodologies are supported. A financial cost which considers a wide remit, for example capital costs, insurance, maintenance or a simple distance based fuel consumption cost can be specified. It is possible to manipulate a driving journey to model a taxi service. The financial attribute for a route utilising a personal car may consider some capital costs, a distance based fuel cost plus associated parking. The same route may be modelled as a taxi route using a known distance/time cost methodology. Thus, the road graph attributes can be levered to simulate varying types of road graph services.

These journey time attributes can also be used in order to make estimations of vehicle volume on the road using a context specific volume delay function (*6*). Emission models that consider GPS informed vehicle speeds, counts and road gradient are also possible.

The road graph also supports modes beyond that of personal car and taxi. The same infrastructure supports walking and cycling modes, with type restrictions. These modes are treated as temporally static, that is to say journey times on a link do not change, thus for a given origin and destination a fixed route will be outputted. Again, a variety of financial cost methodologies can be employed to estimate the attributes of such journeys. Public transport paths often really heavily on road infrastructure for moving between the different graphs. This may be from train to bus, from car to tram and so on.

Public Transport Graph

In contrast to the individual nature of car journeys, public transport services are centrally managed and scheduled. Thus, the quantification of journey times on public transport infrastructure is simpler to harvest but more complex in nature to store as there is the the added complexity of different services with different departure times, routes and journey times. It is only possible to use a public transport service at specified departure times and on specified routes.

In this proof of concept, a weekly schedule was combined with real-time feeds. The weekly schedule sets the planned service schedule for the coming week. The real-time feeds are then used to quantify when, where and how this schedule was changed as a result of planned/unplanned incidents. Unique Association of Train Operating Companies (ATOC) codes are used to match the timetable and real-time feeds to the underlying graph. The unique nature of ATOC codes sidesteps the need to perform geospatial computations. In the case of London, a cost matrix was constructed from published pricing (*32*). In some cities a journey planner outputs pricing information in tandem with a route.

Storing data

Capturing fine spatial and temporal resolution data poses a significant computational challenge. As such the use of a standard GIS database was insufficient and alternative methods were employed. This section will discuss the format of the data and the section on Scaling will explain the reasoning behind this decision.

In order for a decision making methodology to be employed the agent based model must be able to access the relevant spatial and temporal data to present a given agent. This primarily makes use of a light weight data interchange form JavaScript Object Notation (JSON). JSON is built on two structures, first a collection of name/pair values and secondly an ordered list of values (*16*).

The foundation of a graph is the vertex and edge data. Polygon data is used to hold some attributes but is not required to form a graph. For efficiency, the fundamental spatial data is separated from the attributes, resulting in four files types: vertices, vertex attributes, edges and edge attributes. Two example records are presented:

road_vertices:


```
{“group”: 1,
"toid":"osgb4000000031043205",
"point":[508180.748,195333.973],
"index":1}
```


road_vertex_attributes:


```
{"toid":"osgb4000000031043205",
"house_no":6,
“street”: “Hazelbank”,
“locality”:"Croxley Green”,
“administrative_area”: “Rickmansworth”,
“county” : “Hertfordshire”,
“post_code” : “WD3 3EB”,
”country”: “UK",
“elevation” : 58.424}
```


road_edges:


```
{"group”: 1,
“negativeNode":"osgb4000000023183407",
"toid":"osgb4000000023296573",
"term":"Private Road - Restricted Access",
```

"polyline":
[492019.481,156567.076…492126.5,156602],
"positiveNode":"osgb4000000023183409",
"index":1,
"nature":"Single Carriageway"}

roads:

{"group":"Named Road",
"members":
["osgb5000005107792171","osgb4000000023464890"],"toid":"osgb4000000023708569",
"name":"DAPHNE JACKSON ROAD",
"index":1}

Computing Minimum Cost Path

The minimum cost or shortest path route problem is defined as the process of identifying the lowest cost route from an origin to a destination usually in terms of distance, journey time or by a combination of graph edge attributes (such as generalised cost (*9*)). The minimum cost path involves a behavioural decision on human value judgment and this framework seeks to support a range of systems rather than pre-prescribe one.

A large amount of literature exists in the fields of routing and scheduling problems. Significant developments have occurred since Dijkstra presented his path finding algorithm in 1959 (*12*). A range of algorithms (*10, 28*) and software packages may now be taken off the shelf for a range of graph problems. A transport network is generally a directed, weighted, sparse, embedded, explicit and labelled graph. A public transport graph features a further consideration of time constraints (*28*).

There is a large body of academic work which has paired psychological insights to economic analysis in order to better understand human decision making. This is commonly referred to as *behavioural economics*. Within transportation, discrete choice methods have been primarily used to model how an agent makes a decision from a number of discrete alternatives (*5*). It is not the purpose of this framework to pre-define a decision making framework for the modeller but rather to provide a flexible platform where a range of different decision-making methodologies can be supported. As such, this framework focusses on providing relevant input data in a useful and accessible format in order for simple and complex decision making rules to be used.

A general implementation may consider public transportation as bus, rail (all types) and walking, personal vehicle and taxi as road graphs (with different financial attributes) and walking/cycling as road graphs (with type restrictions).

Computing the shortest path for walking and cycling is simple as they are deemed to have no temporal variations. Driving features a large complex graph with temporally dependent edge journey times, directionally restrictions, pricing dependent on mode and agent feedback. Public transport is considerably more complex due to timetabling, variable service routes and a truly multi-modal nature resulting in a series of sub graphs. The walking sub graph on the road network connects rail stations, bus stops and stations have internal subgraphs to connect different platforms and services. Except in the case of public transport, the shortest path may be simply defined as that with the minimum journey time for the modal options model input. For public transport, weight may also be put on the non-time costs of changing services. For example, a time of saving of 2 minutes in exchange of an extra 2 bus changes will usually be deemed undesirable. Each of these lowest cost paths is then presented to the traveller in order for a decision to be computed.

AGENT DECISION MAKING

Agent Logic

Consider a generalised cost methodology where a traveller considers their options in terms of their individual time cost and financial cost. The transport graph presents three modal options, each with their own financial and time cost attributes. Based upon the agent's weight attribution to financial cost (*a*) and time cost (*b*), a generalised cost (*g.c*) may be computed for each option. The agent may then select the lowest generalised cost combination.

$$g.c = a(financial\_cost) + b(time\_cost) \quad (1)$$

It is possible for an agent to give weighted value attribution to time costs dependent on type. For example, one-minute waiting time may be discounted differently in comparison to one-minute in transit as a result of differences in time perception (*13*). A traveller's option is given with type disaggregation and thus a weight *a* may be attributed to the transit time and a weight *b* (where *b* equals a multiple of *a*) may be attributed to the waiting time when computing the generalised cost.

$$g.c = a(train\_time) + b(waiting\_time) + c(financial\_cost) \quad (2)$$

This may be made even more complex, with other factors, such as weather being included. Walking during precipitation can be weighed heavily. This has been investigated by using Met Office NIMROD (*17*) precipitation data as a polygon data input. Consider a situation where two options are posed to a traveller. Option1 features two walking sections with a bus in-between. The second features a taxi with no perceivable walk.

Option1

$$g.c. = a(walking\_time) + b(bus\_time) + c(financial\_cost) \quad (3)$$

Option2

$$g.c = y(taxi\_time) + z(financial\_cost) \quad (4)$$

Where rain is present, the increase in perceived cost of walking (*a*) may outweigh the larger financial cost of the taxi.

With rain:

$$Option1 > Option2 \quad (5)$$

Conversely, without rain the increased cost of a taxi outweighs the now reduced non-monetary cost of walking (*a*).

Without rain:

$$Option1 < Option2 \tag{6}$$

Further complexity could be added by considering a taxi pricing structure which considers supply and demand, such as Uber or Lyft. The increased demand during rain may lead to a pricing tipping point that may even then result in a public transport route with a heavily weighted walking section.

Other metrics that have been investigated include:

- Weight attributions

The discount a traveller attributes to different metrics ay be highly context specific and depend on the environment, role of the traveller and the traveller themselves at the time of the decision. It is possible to adjust this weight with respect to factors such as work vs no work travel, journey distance, travel during weather events etc.

- Spatial information horizon

Travellers may be presented with a spatial limit to their knowledge of the system. This may be used to test the impact of different information strategies for public transport (smart phone versus station focussed) and the efficiencies a centrally coordinated autonomous vehicle fleet could achieve.

- Temporal information horizon

Travellers may be presented with a temporal limit to their knowledge of the system. This may be used to differentiate between an experienced traveller (or smart phone user) and that of a traveller with no historical awareness, such as a tourist. This may manifest itself in a statistical metric, such as the standard deviation of journey times on a given route over a defined time period. This would enable a highly variable road journey or unreliable public transport service to be considered in terms of a risk through the traveller's decision making.

Both temporal and spatial information horizons offer the opportunity to model in-route decision making. Thus, a traveller may decide to alter their route in response to incoming information rather than simply following a one time, static decision for the duration of travel.

Agent Feedback

A key aspect of an ABM is the interactions of individual agents. In order for this to occur there must be feedback between travellers. This is achieved by considering the impact of current travellers on future travellers.

For vehicular travel on the road graph this can be modelled in terms of an increase or decrease in demand. Depending on the ratio of the demand (traffic volume) to supply (road capacity) a travel time impact may be calculated using context specific volume-delay functions (*6*). Thus, a route which is oversubscribed will result in an increase in journey time, potentially leading to a modal change as a result of a change in the input to the decision making process. This is graphically illustrated in Figure 5

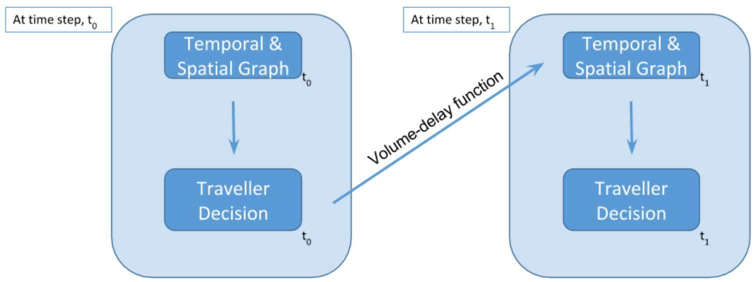

FIGURE 5 Agent decision making feedback via volume-delay function on road network

Feedback on public transport modes requires integration with data feeds that enable the generation and elapsing of travel demand to supply ratios. In order to do this, real-time smart card data (such as Oyster/contactless (*18*)) would be required at a suitable resolution in a timely manner.

SCENARIO TESTING

The use of real-time data sources enables the ABM to inform travellers with realistic real-world data. It is possible to edit these inputs in order to assess the impact of hypothetical changes to the cities infrastructure.

Hard Infrastructure Changes

It is possible to edit the underlying graph to reflect a change in the network. A new link with associated attributes may be added and connected to the existing network. Conversely, links may be removed in order to see the local and global impact of the change. The availability of attributes such as elevation data enables climate change related simulations such as flooding/sea level rises.

Soft Policy Changes

The underlying graphs have a range of associated geospatial tags such as post codes and administrative areas. It is therefore possible to apply policies via a range of attributes. GPS informed journey times allow for the identification of peak congestion allowing for the possibility of dynamic and targeted congestion taxation rather than geographically static taxation.

SCALING

The use of fine resolution data results in high computational demands. The emergence of big-data has led to a shift in how data is stored, processed and analysed.

Decentralized Data Storage

Traditionally spatial data would have been stored in a form of relational database with a specialist setup for geospatial data. Such methods have struggled to scale as they do not inherently support the breaking up of large tasks into smaller sub tasks. In relational databases, references to other rows and tables are indicated by referring to their (primary) key attributes via foreign-key columns. In order to compute the interaction between different elements, joins are computed at query time by matching primary and foreign-keys across many rows of the tables. These operations are compute and memory-intensive and have an exponential cost. Relational databases search all of the data looking for anything that meets the search criteria. The larger the set of data, the longer it takes to find matches, because the database has to examine everything in the collection.

JSON is the data structure of the Web. It's a simple data format that allows programmers to store and communicate sets of values, lists, and key-value mappings across systems. In the present study, the network data are distributed across multiple JSON data files, which allows for a decentralised system for querying and data-processing. The distributed data system allows for easy scalability and load-balancing during computations.

Parallel Computing with Graphs

Relationships are first class citizens in a graph model. A graph is a data-structure that comprises of a set of vertices and a set of edges. Edges represents the path or the relationship between two vertices. There is no need for additional objects to facilitate that relationship. By assembling the simple abstractions of vertices and relationships into connected structures, graph databases enable us to build sophisticated models that map closely to the problem domain.

Data parallelism refers to scenarios in which the same operation is performed concurrently on independent data or elements in a source collection or array across separate resources. In contrast to data-parallel computation, graph-parallel computation derives parallelism by partitioning the graph (dependent) data across processing resources and then resolving dependencies (along edges) through iterative computation and communication (*35*). Graph processing systems apply vertex-centric logic to transform data on a graph and exploit the graph structure to achieve more efficient distributed execution. This form of graph parallel system allows for scalability of multi-modal modelling of big cities. Graph parallel systems are being explored as a means to scale and model real-time big data problems at city-scale.

CONCLUSIONS

A framework has been constructed in such a way so that it can be easily manipulated, can support multiple different classification systems, is self-building, has a fine granularity/resolution and allows for hard or soft manipulation. The ABM allows for the macro and micro impacts of changes to be assessed. Real world empirical data can allow planners to consider how infrastructure actually performs and not how it was designed to perform. The ABM and underlying structures have been built in a distributed fashion in order to facilitate scaling and the use of High Performance Computing (HPC).

The objective of this paper is to describe the general framework of this methodology. A future paper will show an application. This case study will feature how High Speed Rail usage has evolved in the case of HS1 in the UK and what implications this has for low carbon international travel in this region of Europe.

ACKNOWLEDGEMENTS
This research made use of the Directions API by Google and the feeds by TfL. This work was supported by an EPSRC I-case studentship in collaboration with Arup.

NOTE: permission was sought from Google to use the Directions API in this way. Consideration should be made to the terms and conditions of any service.

REFERENCES
1. Balmer, M, Axhausen , K.W, Nagel , K. Agent - Based Demand - Modeling Framework for Large - Scale Microsimulations . Transportation Research Record: Journal of the Transportation Research Board: 125 – 134. 2006
2. Batty, M. Longley, P. Fractal Cities: A Geometry of Form and Function. Academic Press Inc. 1994
3. Batty, M. Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals. The MIT press. 2007
4. Ben-Akiva, M. Bierlaire, M. Discrete choice methods and their applications to short term travel decisions. Handbook of Transportation Science. Volume 23 of the series International Series in Operations Research & Management Science pp 5-33. 1999
5. Ben-Akiva, M. Lerman, S. Discrete Choice Analysis: Theory and Application to Travel. MIT Press. 1985 Available at: https://books.google.co.uk/books?id=oLC6ZYPs9UoC&lpg=PR11&ots=nMfzm-9iHi&dq=discrete%20choice%20transport&lr&pg=PR3# Accessed 07/29/2016. 1985
6. Casey, G. Soga, K. Silva, E. Guthrie, P. Crowd sourced journey times and automated traffic counter volume-delay functions for London. 2016. *In review*
7. Crooks A, Castle C, Batty M. Key Challenges in Agent-Based Modelling for Geo-Spatial Simulation. Geocomputation 2007, National centre for Geocomputation, National University of Ireland, Maynooth, Co. Kildare, Ireland, from 3-5 September 2007.
8. Department for Transport (DfT). Road Traffic Estimates. Methodology Note. 2011. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/230528/annual-methodology-note.pdf Accessed 07/29/2016.
9. Department for Transport (DfT). Transport Analysis Guidance. The Transport Appraisal Process. Transport Analysis Guidance (TAG). 2014. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/431185/webtag-tag-transport-appraisal-process.pdf Accessed 07/28/2016.
10. Desrosiers, J. Dumas, Y. Solomon, M. Soumis, F. Time constrained routing and scheduling. Handbooks in Operations Research and Management Science. Volume 8, Pages 350139. 1995. Available at: http://www.sciencedirect.com/science/article/pii/S0927050705801069 Accessed 07/28/2016.
11. Deville P, Linard C, Martin S, Gilbert M, Stevens F, Gaughan A. E, Blondel V. D Tatem, A.J. Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences, 111, 15888-15893. 2014
12. Dijkstra, E.W. A note on two problems in connexion with graphs. Numerische

Mathematik. 1959

13. Fan, Y. Guthrie, A. Levinson D. Perception of waiting times at transit stops and stations. Transportation Research Part A: Policy and Practice. 2015 Available at: http://nacto.org/wp-content/uploads/2016/02/1_Fan-et-al-Perception-of-Waiting-Time-at-Transit-Stops-and-Stations_2015.pdf Accessed 07/29/2016.
14. Geertman, S. Planning Support Systems (as research instruments. In E. Silva, P. Healey, N. Harris & P. Van den Boek, The Routledge Handbook of Planning Research Methods (1st ed., pp. 322-334). Routledge. 2014.
15. Google. Google Directions API. 2016. Available at: https://developers.google.com/maps/documentation/directions/ Accessed 07/29/16.
16. JSON. JavaScript Object Notation. 2016. Available at: json.org Accessed 07/29/16.
17. Met Office : Met Office Rain Radar Data from the NIMROD System. NCAS British Atmospheric Data Centre. 2003. date of citation.http://catalogue.ceda.ac.uk/uuid/82adec1f896af6169112d09cc1174499
18. Morcency C., Panier T.R, Agard B. Measuring transit use variability with smart-card data. Transport Policy, 14, 193-203. 2007
19. Mtoi, E.T. and Moses, R. Calibration and Evaluation of Link Congestion Functions: Applying Intrinsic Sensitivity of Link Speed as a Practical Consideration to Heterogeneous Facility Types within Urban Network. Journal of Transportation Technologies, 4, 141-149. 2014. http://dx.doi.org/10.4236/jtts.2014.42014
20. Mullick, A. and Ray, A. K. Dynamics of bimodality in vehicular traffic flows. Journal of Applied Nonlinear Dynamics, 3(1), 17-26, 2014. URL arXiv:1205.2314.
21. Office for National Statistics (ONS). 2011 Census. Available at: https://www.ons.gov.uk/census/2011census Accessed 07/29/2016.
22. Ordnance Survey (OS). Integrated Transport Network (ITN) Layer. User guide and technical specification. 2016. Available at: http://digimap.edina.ac.uk/webhelp/os/data_files/os_manuals/osmm_itn_userguide_v1.0.pdf Accessed 07/29/2016.
23. Ratti C, Williams S, Frenchman D, Pulselli R. Mobile landscapes: using location data from cell phones for urban analysis. Environment and Planning B Planning and Design, 33, 727. 2006.
24. Rose, G. Taylor, M. Tisato, P. Estimating travel time functions for urban roads: options and issues. Transportation Planning and technology. http://dx.doi.org/10.1080/03081068908717414. 2007
25. Silva E. A. Waves of complexity. Theory, models, and practice. In: Roo, Gert de, and Elisabete A. Silva 2010. A Planner's Encounter with Complexity, Ashgate Publishers Ltd, Aldershot (UK). pp. 309-331.. ISBN: 978-1-4094-0265-7
26. Silva E. A. Cellular Automata Models and Agent Base Models for urban studies: from pixels, to cells, to Hexa-Dpi's. In: Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban Environment. 2011. Wiley-Blackwell. pp. 323-345. ISBN: 978-0-470-74958-6
27. Skabardonis, A. and Dowling, R. Improved Speed-Flow Relationships for Planning Applications. Transportation Research Record, 1572, 18-23.

http://dx.doi.org/10.3141/1572-03. 1997.

28. Skiena, S. The Algorithm Design Manual, 2nd Edition. Available at: http://www.algorist.com/ 2011. Accessed 04/23/2016.

29. Spiess, H. Technical Note-Conical Volume-Delay Functions. Transportation Science 24(2):153-158. http:// dx.doi.org/10.1287/trsc.24.2.153 1990.

30. Transport for London (TfL). Traffic Modelling Guidelines. Version 3.0. 2010 Available at: http://content.tfl.gov.uk/traffic-modelling-guidelines.pdf Accessed 05/02/2016.

31. Transport for London (TfL).. Open Data Users - Live feeds. 2016 Available at: http://www.tfl.gov.uk/info-for/open-data-users/

32. Transport for London (TfL). Fares & Payments. 2016. Available at: https://tfl.gov.uk/fares-and-payments/ Accessed 07/29/16.

33. Wang F, Xu Y. Estimating O-D travel time matrix by Google Maps API: implementation, advantages and implications. Annals of GIS, 17:4 199-209. 2011

34. Wu, N., Silva, E.A. Artificial intelligence solutions for Urban Land Dynamics: A review. Journal of Planning Literature. 24:246-265. 2010

35. Xin, R. S., Crankshaw, D., Dave, A., Gonzalez, J. E., Franklin, M. J., & Stoica, I. GraphX: Unifying data-parallel and graph-parallel analytics. arXiv preprint arXiv:1402.2394. 2014.