MATCHING PUBLIC TRANSPORT SERVICE AND PATRONAGE DATA TO OPENSTREETMAP

Mark Dimond
Integrated Transport Planning Ltd.

David Brenig-Jones
Integrated Transport Planning Ltd.

SUMMARY

This paper outlines a method that enables geographic data from very precise sources (in particular satellite positioning) to be used for different transport modelling tasks. It explains the importance of 'map matching' precise data to road networks, and examines how OpenStreetMap can be used as a flexible source of freely available road network data. We will describe a successful project to collate on-vehicle patronage surveys using road network data, and outline the benefits of aggregating data in this way.

1. INTRODUCTION

The importance of geography to modelling of transport networks can hardly be overstated. The last ten years of advances in mobile positioning and communications have brought huge benefit to many fields, but are not yet exploited to their fullest extent in understanding transport networks. In particular, data collected through the usage of Global Positioning System (GPS) satellite technology can be both precise and repeatable enough to form a comprehensive evidence base through surveying or sensors. This paper will describe a method for more readily incorporating data from GPS into transport models of both supply and demand, exploiting it to produce a detailed understanding of real world transport.

Bus and tram networks are often considered in terms of supply (availability of services) and different forms of patronage and demand (use by passengers). The data mining approach we have developed can be applied to both aspects. Surveys of public transport supply are often used in developing countries where timetable infrastructure may not be complete; patronage surveys can be used both in the UK and internationally to understand the numbers of passengers using transport services at different points along routes. The importance of the process described here is in the automation of large parts of the processing of these surveys, collating supply and demand data across whole cities.

The advantage of this approach is that it greatly reduces the requirement for manual postprocessing of survey data, freeing up resources for analysis tasks. It also permits a range of new types of analysis, examined in detail below. Though at present surveys of supply and demand are collected manually, our data mining technique also forms the first step towards use of sensors for collecting supply and demand data. Ultimately a range of data sources, including vehicle telemetry, electronic ticketing systems, and CCTV data could all be used to understand transport network supply and demand. With this automation of data collection, the costs and timescales required for understanding changes in city transport networks would greatly reduce, opening the door to real time monitoring at a street-link level of detail.

2. BACKGROUND

Data sources for transport network modelling vary. In the UK, the TransXChange format captures availability of bus and rail services, with detailed route data and stopping schedules (Department for Transport, 2014). Internationally, the General Transit Feed Specification (GTFS) uses a simple text-based format to capture similar data about transport for cities and regions. Both of these sources have potential to produce a rich understanding of network supply. Accessibility analysis can examine different sites around cities and relate them to population data or access to jobs or services, in some cases performing powerful analysis with an interface as simple as dropping a point on a map and visualising areas within set travel times. This valuable analysis can use Basemap's TRACC product (Basemap, 2016) in the UK or Conveyal tools internationally.

2.1 OpenStreetMap

Applications that exploit GTFS internationally can draw from other open data sources to support their functionality. The most prominent of these is OpenStreetMap (OSM), a worldwide, crowdsourced mapping effort run by volunteers and open to anybody to edit. OSM and its related suite of tools allow users to log in and digitise map data with minimal training (Goodchild, 2007), and its popularity means quality is maximised through collaboration. It is particularly powerful in developing countries, where national mapping agencies may not exist or may not be well funded, but also works well in developed nations,

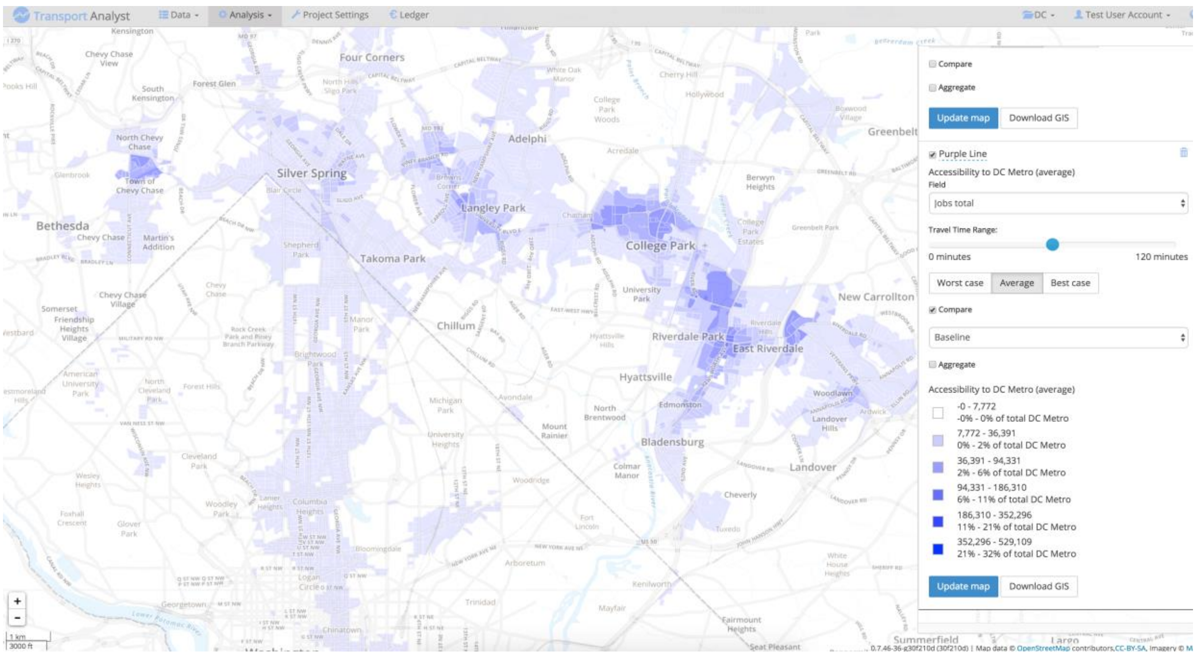

Figure 1. OpenStreetMap data is important for interactive accessibility analysis.

where data captured by researchers and enthusiasts provide very good coverage in most regions.

In some circumstances, OSM mapping has been found to rival the quality of that produced by government-funded mapping agencies and by private providers (Ludwig *et al*, 2011). This allows building of many different types of transport data services with detailed mapping which is internationally available at low cost. For example, the Open Source Routing Machine[1] project provides a directions service across any part of the OSM 'planet'. The OpenTripPlanner[2] project aims to develop a full-featured multimodal journey planning application that incorporates street network restrictions and accounts for availability of transport services through GTFS data. Other possibilities include the detailed public transport accessibility analysis (described above) such as that offered by Conveyal, an example of which is shown in Figure 1.

The biggest benefit of OSM to transport planning is the very flexible vector data format in which mapping is captured. This provides the ability to extract specific parts of the data format required for different tasks. Of particular use is the ability to access road network data in this fashion, and detailed capture of constraints such as pedestrianisation or link classification. In Figure 2, the car-accessible road network around Nottingham city centre can be seen as an example. OpenStreetMap therefore represents a very valuable data source for transport planning applications. It is free and available to a good standard internationally, and where volunteer-supplied data is not sufficient, users can update mapping themselves.

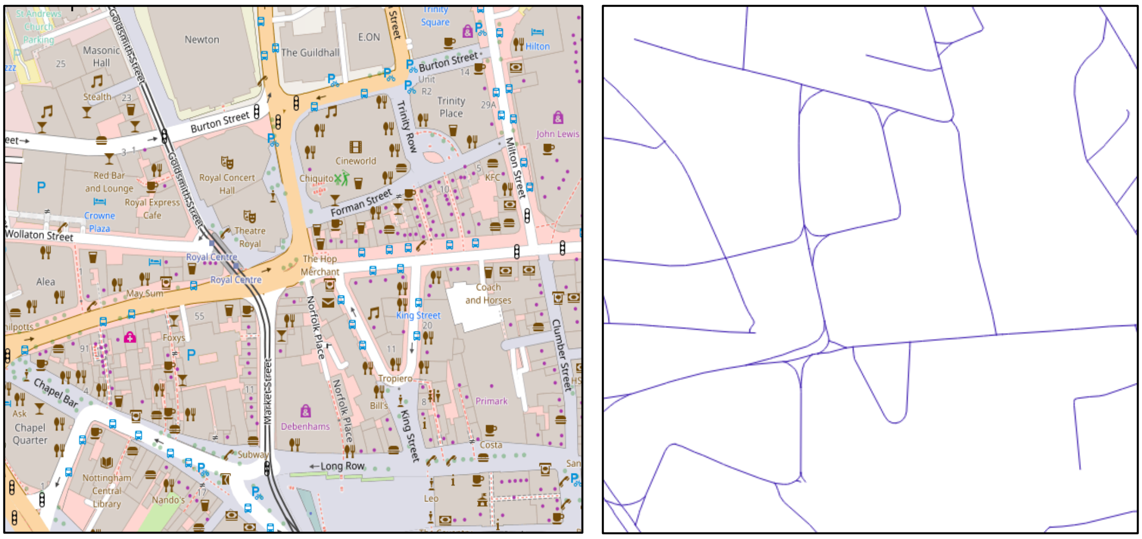

Figure 2. Visualised data from OpenStreetMap for a portion of Nottingham (left), and the vector data extracted for the routable road network (right).

Using detailed data sources such as OSM can complement different types of approach, for example matrix-based origin destination (OD) travel modelling. In some circumstances, the ability to calculate routes over a detailed road network dataset can enhance these methods. The Propensity to Cycle Tool (Lovelace *et al*, 2017), for example, includes approximations of daily routes taken by cycle commuters, using the CycleStreets tool to apportion census data to the network.

2.2 Map Matching

To draw the greatest possible benefit from open data sources, we have developed an approach to corroborate different types of network data, such as service runnings and patronage surveys, with underlying network information drawn from sources such as OpenStreetMap. This is necessary as this type of data is usually collected through human digitisation or GPS surveys. Both of these are subject to error and do not typically capture 'topology' data (linking to specific elements of a transport network or other infrastructure).

This task, known as *map matching*, has been well studied in GIS research in the past. Quddus *et al* (2007) presented an initial review. The seminal work of Newson and Krumm (2010) at Microsoft used a machine-learning approach (known as a hidden Markov Model) to filter GPS noise and estimate the correct road segment for each point despite error in observations. Finally, an updated approach presented by Wei *et al* (2013) refines this technique and reviews others, and presents use of the Fréchet distance to minimise error, further improving performance. These approaches gradually improved handling of both noise (high positional error) and sparseness (low sampling rate) in traces of GPS data.

We use a simplified approach to map matching using street network routing, technical details of which were presented initially in a previous work (Dimond *et al*, 2016). This tool uses open-source database and routing software[3] to estimate routes between GTFS stop timings, in a similar fashion to Griffin *et al* (2011). However, we also presented a web tool which allows users to interactively correct map matching using rapid selection or moving of estimated routes, ensuring map-matched GTFS data is of high quality. The following section describes this process in detail, before examining the utility of a map-matched supply network in understanding patronage survey data.

3. MATCHING SERVICE DATA

Introduced above, OpenStreetMap represents an excellent dataset for examining the geography of regional road and transport networks. However, to maximise its value for transport planning, it must be combined with public transport supply in formats such as GTFS. Our previous work (Dimond *et al*, 2016) proposed a technical approach for this using routing and a correction interface. This operates by parsing each *route*, *trip*, and *stop time* as contained within every GTFS feed, capturing the public transport services for a city:

1. The OSM road network for a city is parsed into network 'segments', sections of highway that represent unbroken links between junctions.
2. In GTFS, service trips are captured as sequences of timed stops. A data mining process estimates which road segments are required to navigate a city or region's roads. This uses the pgRouting street directions tool[3] to estimate valid routes between points, accounting for constraints such as directionality of links, turn restrictions, and road classifications.

3. An estimated set of route 'alignments' are recorded for each stop within a trip in a GTFS feed, capturing the use of the road network by that service.
4. The estimated street routing is reviewed by a knowledgeable user, with a web interface allowing visualisation and rapid correction of the captured routes.

The web visualisation required for reviewing estimated alignments represents a significant component of our previous work (see Figure 3). An interactive web interface displays the estimated road links, and a simple click-select interface allows the user to rapidly make corrections to the route. Upon completion of this process for all trips, the GTFS feed is fully map matched to the road network, allowing powerful analysis of the transport supply network relative to the roads. This 'alignment' data model links sections of trips with the road network. Queries of these alignments can include checking service coverage, identification of delay-prone route sections, and feasibility of interventions such as re-routing of services.

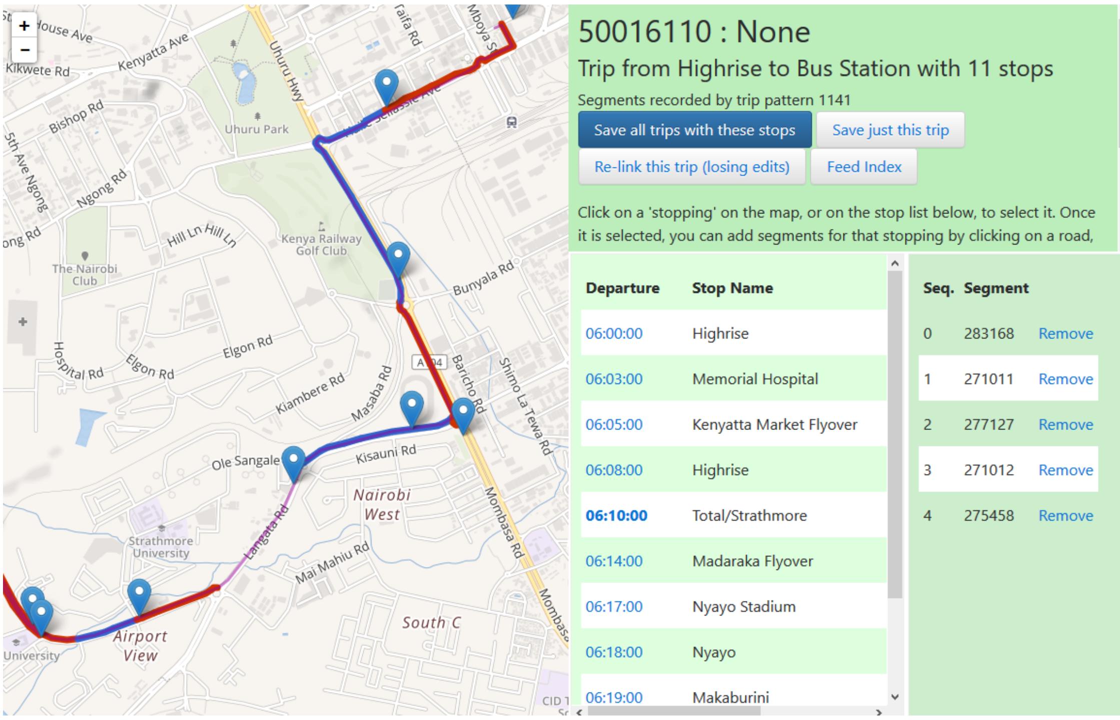

Figure 3. Manually correcting the estimated route on OSM for a public transport trip.

A further advantage of a fully map-matched public transport network is the higher quality of data then available: using exact road alignments represents network data that is of high enough standard for wider publication in route maps or network information (unlike that produced through GPS or digitisation). A map-matched public transport network is also an important first step in exploiting road network data for understanding of public transport patronage, explored further in the following section.

4. MATCHING PATRONAGE DATA

An important part of improving public transport provision is understanding the levels of use experienced by existing services. In the past, this may have required paper-based counting of passengers on vehicles or broad, high-level analysis of ticket sales. However due to the increasing availability of electronic ticketing data, onboard WiFi connection information, and mobile telecommunications, better approaches are becoming available.

ITP has pioneered the use of app-based manual surveys to capture detailed public transport patronage data using the TransitWand app. This tool, developed in partnership with Conveyal, allows recording of GPS traces of vehicle routes by an on-vehicle surveyor, who can also use a counter interface to track the boarding and alighting of passengers at specific stops. Postprocessing of this data allows calculation of a wide range of important statistics for each segment in the road graph: average vehicle loading, traffic speed, most frequently used stops, etc. Patronage data can be aggregated across many vehicles and filtered to a specific time of day.

Recently developed technology partially automates this postprocessing, greatly improving upon a workflow that hitherto required significant GIS expertise. Recorded GPS traces from TransitWand do not require 'true' map matching against the road network, if the public transport route surveyed has been captured in a geographic format. For our work, we take bus and tram routes recorded in GTFS, link them to a road network as described in the previous section, and simply record patronage against the roads used by the captured network.

Patronage data recorded from manual surveys in this fashion are marked as having occurred at particular points on road segments, and at specific dates and times. Appropriate aggregation can then be used to calculate road segment-level statistics for:

- passenger and vehicle load
- timetable performance
- traffic speeds

An interactive web interface (see Figure 4) allows the rapid visualisation of the most common of these queries for a set of collected TransitWand data, giving the user the chance to explore the results of collected surveys. It is also possible to produce 'snapshot' data of speeds or passenger loads for every road segment within a city or region, showing the most busy or congested areas in a GIS-supported format, subject to sufficient data having been captured.

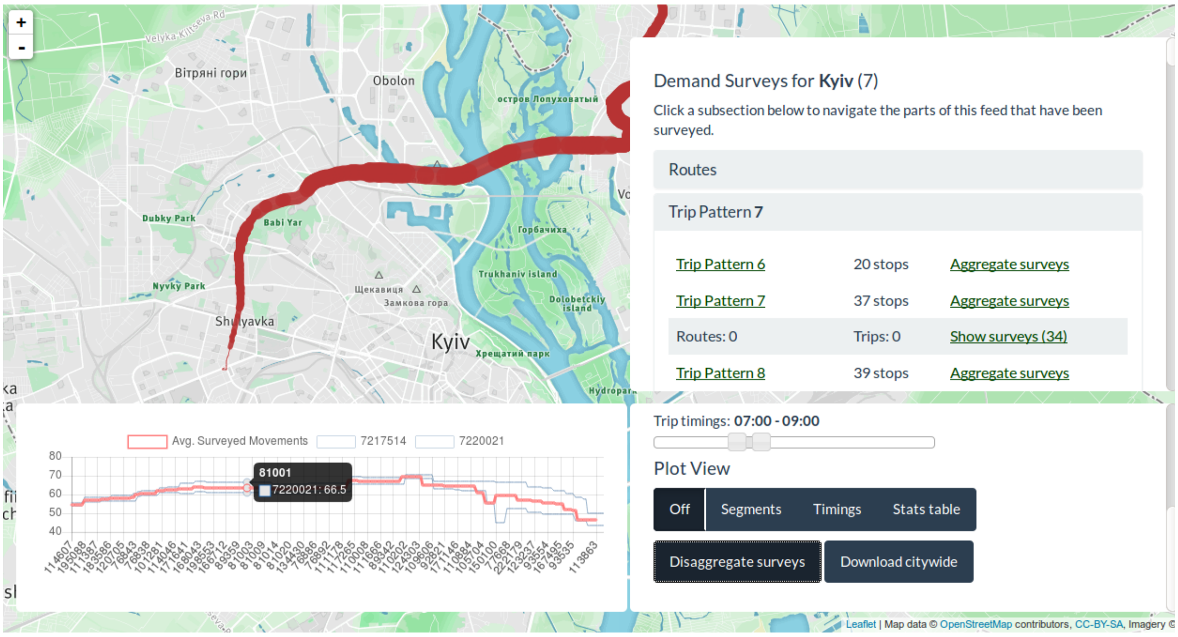

Figure 4. Interactive exploration of demand data using road network alignments

5. CHALLENGES

Matching public transport supply data to OpenStreetMap proves powerful and flexible. However, the use of routing to estimate the exact streets taken by each service means that a requirement remains to validate output. Our previous work estimated approximately 83% correctness of selected streets (Dimond *et al*, 2016, p4); it is expected that this might be greatly improved in future work by using a probabilistic map matcher such as the hidden Markov Model used by Newson and Krumm (2009). A requirement for validation will remain, as the probabilistic approach nevertheless relies upon routing to some degree and in many cases transport services do not necessarily follow shortest routes.

The current requirement for manual survey of transport routes to determine levels of patronage is a significant barrier to wider adoption of open road network data. At present, to record levels of bus or tram use, surveyors must be recruited and trained to travel on vehicles counting the boarding and alighting of every passenger. Though the data collected is valuable, the labour requirement represents a high cost, and a barrier to repeatable or large-scale surveying. Our data model has been developed to be flexible for different data sources, in the expectation that future patronage assessment might be based on electronic ticketing, WiFi connectivity, and vehicle telemetry. Ongoing work is investigating the requirements of using these ‘incidental’ sources of patronage data for better public transport intelligence.

Finally, the compatibility of transport supply data standards presents a minor challenge for development of our OpenStreetMap-based tools. Many of ITP’s transport network modelling clients are based outside the UK, with a large number in developing countries. In such cases, availability of transport routes is often captured in the lightweight GTFS format.

However, the UK's strong history of open data means that it uses the more detailed TransXChange specification to capture its service data. As a result, many GTFS data tools cannot be directly deployed here. However, this can be mitigated by creating a simplified GTFS network for specific UK transport regions or corridors, or through the use of conversion tools. Future work will allow the reading of TransXChange data directly into our map matching and visualisation tools.

6. CONCLUSIONS

This paper presents an introduction to the use of map matching for exploitation of topology in transport planning. It briefly outlines the basis of an algorithm for performing this map-matching, and explains the methodology for using matched routes for understanding both supply and (in the form of patronage) demand.

Future work will expand the routing-based matching algorithm to account for different modes of transport within a region, seamlessly reading the type of service (tram, bus, or metro) from a format such as GTFS and allocating to a different network as appropriate. The interface for querying alignment of services and patronage data will also be updated as newer data sources become available, bringing the benefits of increased digitisation in ticketing and payment to bear upon planning regional transport networks.

ACKNOWLEDGEMENTS

The Demand Visualisation tool was supported in part by a Department for Transport Technology in Transport Research Innovation Grant ('TTRIG')

NOTES

1. Open Source Routing Machine, modern C++ routing engine for shortest paths in road networks. http://project-osrm.org/
2. OpenTripPlanner, multimodal trip planning and analysis http://www.opentripplanner.org/
3. pgRouting, open source routing library, http://pgrouting.org/

REFERENCES

Basemap Ltd. (2016). TRACC v1.2.2 Released, product announcement, Surrey, 2016.

Department for Transport (2014), TransXChange Schema Guide 2.5 v59, issued schema documentation, 2009-2014.

M. Dimond, N. Taylor, and R. Houghton (2016). Estimating and Editing Transit Topology over the road graph using supply data feeds; to appear in proc. of Assoc. of Geospatial Information Laboratories Europe (AGILE) 2016, Helsinki.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography, GeoJournal 69:4, pp211-221.

Griffin, T., Huang, Y., and Seals, S. (2011) Routing-based map matching for extracting routes from GPS trajectories. COM.Geo 2011

Lovelace, R., Goodman, A., Aldred, R., Berkoff, N., Abbas, A., Woodcock, J. (2017). The Propensity to Cycle Tool: An open source online system for sustainable transport planning. Journal of Transport and Land Use. 10:1, pp505–528.

Ludwig, I., Voss, A., and Krause-Traudes, M. (2011) A Comparison of the Street Networks of Navteq and OSM in Germany, Springer Lecture Notes in Geoinformation and Cartography.

Newson, P. and Krumm, J. (2009) Hidden Markov map matching through noise and sparseness ACM SIGSPATIAL 2009

Quddus, M., Ochieng, W., and Noland, R. (2007) Current map-matching algorithms for transport applications: State of the art and future research directions. Transportation Research Part C.

Wei, H., Wang, Y., Forman, G., and Zhu, Y. (2013) Map matching: Comparison of approaches using sparse and noisy data ACM SIGSPATIAL 2013