New Routes on Old Railways: Increasing Rail's Mode Share Within the Constraints of the Existing Railway Network

Simon P Blainey[a]*, John Armstrong[a], Andrew S J Smith[b], John M Preston[a]

[a]Transportation Research Group, Faculty of Engineering & The Environment, University of Southampton, Boldrewood Campus, Burgess Road, Southampton, SO16 7QF, United Kingdom

[b]Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, United Kingdom

* Corresponding author. Tel.: +44 (0)23 8059 2834; fax: +44 (0)23 8059 3152; email: S.P.Blainey@soton.ac.uk

ABSTRACT

This paper describes an integrated methodology for identifying potential 'quick wins' for mode shift from road to passenger rail transport. Firstly, a procedure for analysing rail's relative competitiveness in the market for passenger transport between large urban areas is developed and then applied to a UK case study. The purpose of such analysis is to allow the identification of flows where rail is currently relatively uncompetitive (in terms of journey time in particular) and to assess the reasons for this poor performance, so that the issues which suppress rail use may be addressed. In parallel, a framework, methodology and tool for the assessment of existing and potential capacity (trains, seats, TEUs, etc.) is developed for both passenger and freight traffic, to identify and address network constraints. An illustrative example of the use of these demand and capacity assessment tools is then presented, with the tools used to identify and evaluate flows where rail demand is suppressed by poor service quality and where spare capacity exists which would allow the passenger rail service to be improved without requiring significant investments in infrastructure. The effects of such improvements on demand are predicted, and the cost implications of operating such additional services are discussed. The analysis suggests that there may be significant potential for increasing rail's mode share by providing additional inter-urban services where rail currently offers an inferior service.

Keywords:

Rail capacity; Rail demand; Rail costs; Modal shift

1 INTRODUCTION

The current network of passenger rail services in the majority of developed countries is largely the result of incremental historical development with its origins in the pattern of demand for travel in the 19th and early 20th centuries. Patterns of rail usage are inevitably and to a large extent shaped by this service pattern, which means that rail will tend to have a greater share of the travel market for flows which have traditionally been well-served by rail services and where people have therefore been able to base their lifestyle choices (such as residential location) around regular rail use. This is accentuated by the significant role which habit plays in determining travel behaviour (Thøgersen, 2006), which may mean that people in areas with a greater 'tradition' or 'culture' of rail use will be more likely to use the train. Habit can also play a significant role in transport planning, for example with regard to the structure of rail services and timetables. Travel patterns and the relative importance of different travel flows have changed markedly since road travel overtook rail as the dominant travel mode in the first part of the 20th century, as have land use patterns, but the structure of the rail services provided has often failed to mirror these changes. This lack of innovation poses problems given the wider context in which the industry operates, as it will almost certainly be necessary to achieve a significant level of mode shift from road to rail to meet the challenges posed by increasing road congestion and the urgent need to cut carbon emissions. This will require rail to capture a large number of trips from the car on flows where rail currently has only a minimal market share.

This paper describes the development of a methodology which analyses current travel options, volumes and mode share across a matrix of flows, assesses the availability of network capacity for additional services, and compares the costs of introducing such services with the revenue generated from additional passengers. This then allows the identification of any flows where additional services could be justified on the basis that the additional revenue generated would exceed the costs incurred, and also indirectly of flows where the subsidy per passenger kilometre of a new service would be lower than the subidy per passenger kilometre required to operate some existing services. By initiating a comprehensive analysis of inter-urban travel by all modes, this methodology has the potential to help rail service planning break free from the constraints of 'historical' rail markets and service patterns, and generate significant mode shift from road (and, to a lesser extent, air) to rail for target flows. It forms part of a wider body of research under the heading 'Factor 20' (see the Acknowledgements section below), with the aim of increasing rail passenger numbers ten-fold through mode shift from road and air, while at the same time reducing rail's carbon emissions by 50%, thus increasing rail traffic per unit of carbon emitted by a factor of 20.

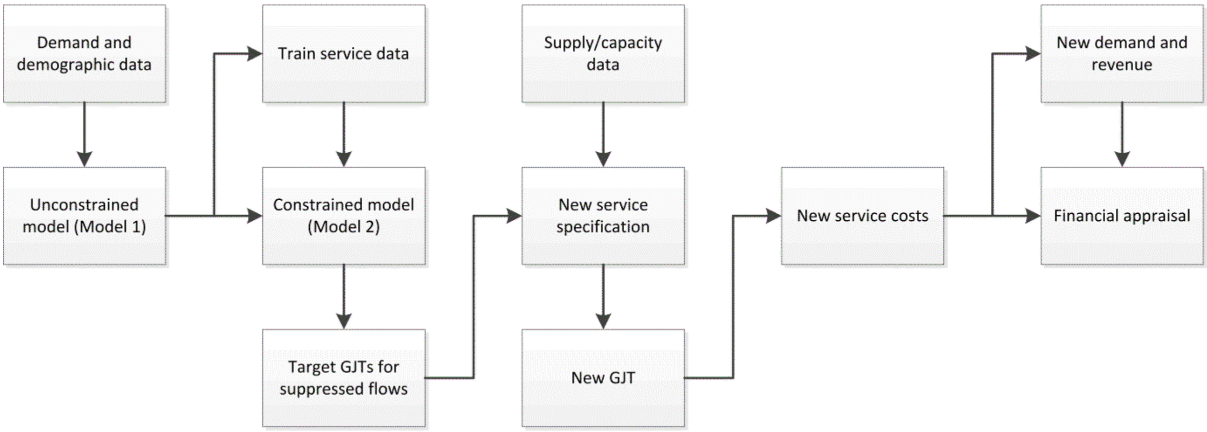
2 BACKGROUND

In Great Britain frequencies on many rail routes are at the highest ever levels, but in many cases the spatial pattern of rail services differs little from that which was offered to travellers in the mid-20th century. Even though passenger journeys grew by 98.9% between 1994/95 and 2011/12 (Department for Transport, 2012b), there are still areas where rail services are under-developed despite this wider story of growth. Part of the reason for this may be that there is relatively little incentive for the privatised Train Operating Companies (TOCs) to engage in innovative route planning, with the short length of franchises inhibiting long-term strategy (McNulty, 2011), a point backed up by the innovation displayed by Chiltern Railways, holders of a 20 year franchise. The fact that the majority of rail services in the UK are subsidised is also likely to limit innovation (on average, operators received subsidy of 8.1 pence per passenger mile in 2010-11 (Department for Transport, 2011)), as TOCs are unlikely to institute new services on their own account unless they are profitable, whereas there might be potential for additional services which, while requiring subsidy, would need less subsidy to operate than many existing services (average subsidy in 2010-11 varied from -0.9 to 34.0 pence per passenger mile between TOCs). UK TOCs are in the main maximising profit subject to a series of constraints set in their franchising agreements, a situation which is very unlikely to lead to a welfare-maximising fare service combination. 'Open access' operators have begun to serve a small number of non-traditional rail flows in recent years, but there are a number of barriers to entry in the rail passenger market, and the withdrawal of the Wrexham, Shropshire and Marylebone open access services (one of only three such operations in the UK) in early 2011 highlights the difficulties faced even by high quality operations.

While these problems are in some cases unique to the UK, even where they do not apply the scope for innovation and expansion in service provision is in any case often limited by infrastructural constraints. This may be because infrastructure construction has not kept pace with changing land use and travel patterns, or may reflect a reduction in network capacity and flexibility where declining traffic and improved operational efficiency in areas such as signalling meant that a constant or declining number of trains could be operated on a reduced infrastructure. This can make it difficult for the railway to accommodate and adapt to expanding levels and/or changing patterns of traffic, and the difficulties and costs involved in reinstating or expanding infrastructure emphasise the need to extract additional capacity from the existing system by whatever means possible.

It could be argued that the lack of innovation in route planning is of little consequence, for as long as the rail industry continues to serve its 'traditional' markets well, it will be fulfilling its purpose as a transport mode, even if it does not provide a realistic service option on other flows. However, as stated above this disregards the wider context within which the industry operates, which means that there are compelling arguments for implementing policies aimed at increasing rail's mode share. While in many cases improving provision on flows where the rail service is currently poor or non-existent will involve major infrastructure investment, there may be some flows where the current rail service provision is poor but where capacity exists to make significant improvements without the need for major infrastructure investment, offering the possibility of 'quick wins' in terms of modal shift. In order to investigate whether or not this is the case, a methodology has been developed to analyse current travel options, volumes and mode share across a matrix of flows, with the aim of identifying the flows with the greatest potential for expansion and development. Alongside this demand side analysis, a capacity analysis framework was developed to allow the identification of flows where capacity for improved services exists. The results from these analyses are then brought together with the estimated costs of new services to assess the likely costs and benefits of improving the rail service offer. The structure of the methodology is shown in Figure 1, and the remainder of this paper describes an initial implementation of the methodology using a British case study.

Fig. 1 Structure of methodology for improving the demand-capacity balance



2 DEMAND ANALYSIS

2.1 Choice of Case Study

The first stage in the study was to select a case study dataset of urban areas for analysis. Ten sample urban areas in Central and Northern England were selected from a list of the 100 largest British urban areas to form a case study for the initial implementation of the methodology, giving a matrix of 100 flows. These areas were selected so that the sample included a range of urban area sizes and locations, and flows with rail services ranging from excellent to virtually non-existent, while at the same time restricting the amount of the rail network included to keep the scope of the capacity modelling manageable. They are mapped in Figure 2 along with the rail network which links them.

2.2 Data Collection and Preparation

Classification and definition of rail flows between urban areas was reasonably straightforward, as the relevant flows were assumed to be those between the major station (or, in a few cases, stations) in each urban area.

Fig. 2 Urban Areas Included in Case Study Superimposed on Rail Network

Car travel was more problematic, as the door-to-door nature of such travel means that there will be no single central origin or destination for car trips in any urban area, and without comprehensive long-distance travel survey data (which do not exist for the UK) it is not possible to establish exactly where people travel from and to. However, the ‘door-to-core network’ and ‘core network-to-door’ elements of ‘door-to-door’ inter-urban trips will form only a small proportion of the total distance travelled in most cases, and therefore for the purposes of this study all car journeys were assumed to be made to and from the main railway station, to give comparability with other modes. While this is analogous to the use of zone ‘centroid connectors’ in a range of transport modelling applications, and should give a reasonable estimate of the mean travel distance and time for drivers between particular urban areas, it may mean that results are slightly biased in favour of rail for flows to and from urban areas where the main railway station is a long way from the town or city centre (for example Rochdale). A similar (albeit less severe) problem may exist in some cases for bus and coach travel, where there will often be a number of stops within the urban area, and in this case the central coach/bus station, where one exists, was used as the origin/destination, with the stop closest to the main railway station used in all other cases.

Total travel volumes between urban areas will be influenced by the characteristics of the journeys between them, and these characteristics will almost entirely determine mode shares on particular flows. The most fundamental characteristic of flows between urban areas is the distance covered, and this spatial separation was defined based on road distance to account for geographical barriers to access. To provide comparability between modes, speeds for different modes were all calculated based on this distance rather than the distance actually travelled. Spatial data on the road and rail networks were obtained in Ordnance Survey 'Meridian' vector format via the OS OpenData agreement, giving a GIS-based representation of these transport systems.

A number of different variables characterising the rail service for each urban pair were collected from a range of internet sources, including average rail journey time, interchange time, service frequency and hours of operation, and rail fare. It would be desirable to include rail crowding levels in the modelling, but while data on approximate crowding levels is available in Network Rail's Route Utilisation Strategies (Network Rail, 2007, 2009, 2010), this was not the case for all flows, and a comprehensive dataset could not be constructed. A similar situation applies with regard to bus and coach services, with data on journey times, service frequencies and hours of operation again collected from internet journey planners. Service frequency and hours of operation are not an issue for car travel, but journey times are still crucial, and four collection methods were tested for this study. These comprised estimated times from three internet journey planners, specifically Google Maps (maps.google.co.uk), AA Route Planner (www.theaa.com/route-planner) and Transport Direct (www.transportdirect.info), and times estimated using ArcGIS based on Ordnance Survey Meridian road network data. While there were a number of disparities between journey times estimated using the different methods these predominantly affected short trips, and the choice of method should not prove crucial for the inter-urban trips considered here.

Probably the most comprehensive dataset on UK travel volumes is provided by the 2001 census travel to work data. However, this only covers commute trips, and a disclosure control procedure known as the ‘Small Cell Adjustment Methodology’ means that the numbers of people recorded as travelling by minority modes (which in many cases includes rail) or on low volume flows can be unreliable. Because of this, a rail demand dataset controlled by the UK Department for Transport (DfT) called The Oxera Arup Dataset (TOAD) was used to depict travel volumes. This dataset comprises rail demand data for over 23,000 flows over some or all of the period 1990-2008, disaggregated by ticket type, and also data on a number of associated explanatory variables. Because the dataset does not include flows where total revenue was less than £10,000 in 2005/6 or where there were ±100% year to year changes in demand (Arup & Oxera, 2010), data were only available for 67 of the possible total of 90 flows between the matrix of urban areas. While TOAD only contains travel volumes for a single mode, it should still enable the identification of flows where rail performs poorly in terms of passenger numbers by comparing predicted rail demand based on urban area characteristics and proximity with observations of actual rail use. Similar electronic ticket sales data should be available in most countries which operate centralised ticketing systems, and could be used to conduct equivalent analysis elsewhere.

2.3 Analysis of Relative Rail Demand Levels

An extensive body of research exists on modelling rail demand both in the UK and elsewhere (see for example Blainey & Preston, 2013; Brons et al., 2009; Lane et al., 2006;

Wardman et al., 2007). The majority of the UK-based evidence is brought together in the Passenger Demand Forecasting Handbook (PDFH) (Association of Train Operating Companies, 2009), which recommends the use of aggregate regression models for predicting rail demand at the flow level. While this general approach was used as a basis for the modelling undertaken here, along with previous rail demand forecasting work undertaken by the study team (Blainey & Preston, 2009), the requirement for forecasts which isolated the effects of service quality meant that a bespoke methodology was required.

Analysis focused first on establishing the expected level of rail demand for particular flows, given the characteristics of the urban areas at either end of the flow and the distance between them. Other journey characteristics were not considered at this stage, since in order to establish rail's relative performance in attracting demand, it is necessary to estimate the likely level of demand if the rail service (and the attractiveness of other modes) were the same for all flows. Two means of representing urban areas were tested, using dummy variables and using various combinations of socioeconomic variables such as population and employment levels. The latter did not improve model fit and increased the model complexity, and the dummy variable model (given by equation (1)) was therefore chosen as the preferred model. The results of the model calibration are summarised in Table 1, which shows that the distance parameter was strongly significant.

$$T_{rij} = \alpha\left(\prod_{i} {\beta_i}^{O_i}\right)\left(\prod_{j} {\gamma_j}^{D_j}\right)K_{ij}^{\rho} \quad (1)$$

Where:

T_{rij} is the number of rail trips from urban area *i* to urban area *j* in 2008 as recorded by TOAD

O_i is a dummy variable with the value 1 if the origin urban area is area *i*, and 0 otherwise

D_i is a dummy variable with the value 1 if the destination urban area is area *j*, and 0 otherwise

K_{ij} is the road distance from urban area *i* to destination urban area *j* in km *α, β, γ* and *ρ* are parameters determined during calibration

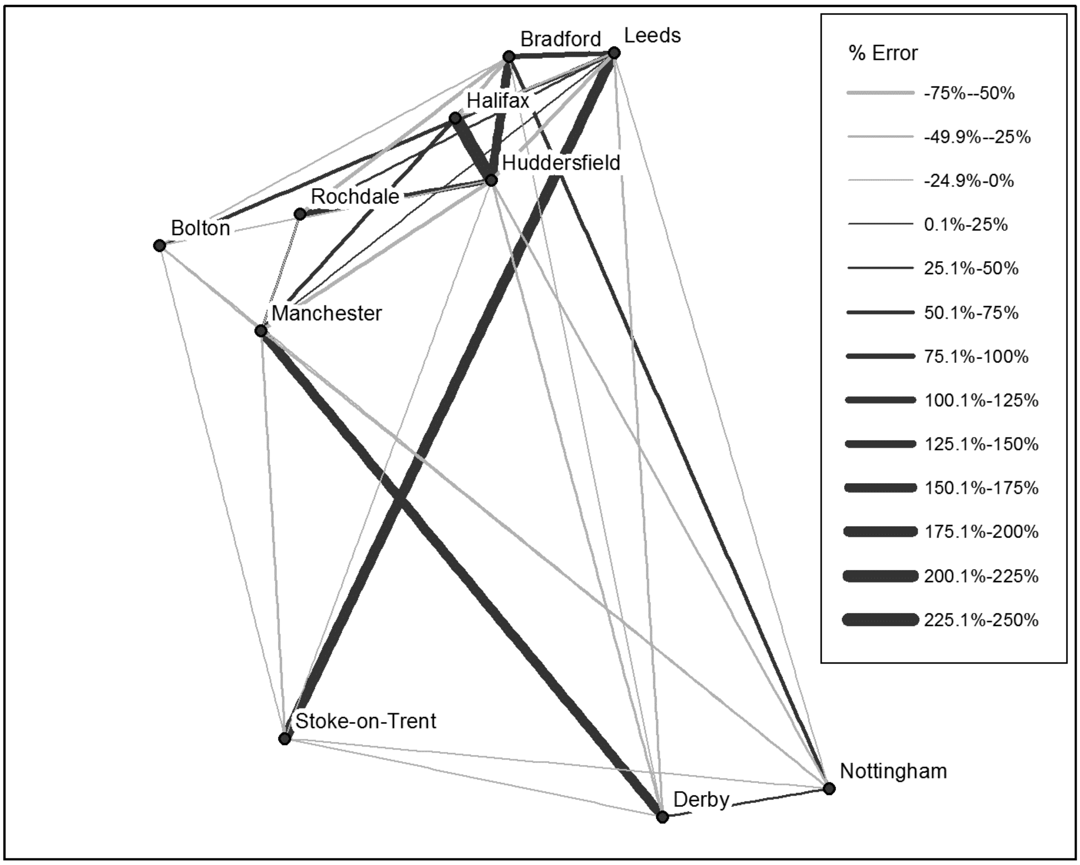
Table 1: Summarised Results From Calibration of Model (1)

	Value	t stat
Intercept	15.261	22.869
Road distance (ρ) parameter	-2.045	-14.731
R_{adj}^2	0.895	

The percentage prediction errors from Model (1) are mapped in Figure 3, with dark grey lines showing overpredictions, and light grey lines showing underpredictions. This illustrates that, on several flows, fewer rail trips are made than might be expected, given the distance between the urban areas and their trip generation/attraction potential.

In order to establish the effect that rail service levels have on demand, a further range of variables were introduced into the models. These included rail service frequency, rail service operating hours, rail fare per kilometre (anytime and cheapest available), rail speed (based on rail travel time), and rail speed based on generalised journey time (GJT). These GJT estimates were obtained from TOAD, and are calculated by combining the total station to station journey time with a service interval penalty and the sum of the interchange penalties, as outlined in the PDFH. Because many of these variables cover essentially the same characteristics of the underlying data, a range of models were tested to determine the best model form. Road speed variables were then added to account for the effects of competition from car travel, with several definitions of road speed tested. The socio-economic variables representing urban areas were also tested again, to confirm that they did not significantly affect model fit. The best results were given by Model (2), and these calibration results are summarised in Table 2, which shows that incorporating additional variables led to an improvement in model fit, and that the speed of the rail journey has a greater impact on the level of rail demand then the speed of the car journey.

Fig. 3 Percentage Prediction Errors From Model (1)



$$T_{rij} = \alpha\left(\prod_i \beta_i{}^{O_i}\right)\left(\prod_j \gamma_j{}^{D_j}\right)K_{ij}^{\rho}RfC_{ij}^{\nu}RsGJT_{ij}^{\xi}CsI_{ij}^{\zeta} \quad (2)$$

Where:

RfC_{ij} is the cheapest rail fare per km of road distance from urban area *i* to urban area *j*

$RsGJT_{ij}$ is the mean GJT-based rail journey speed (the road distance divided by the GJT from TOAD) from urban area *i* to urban area *j*

CsI_{ij} is the mean of the off-peak mean road journey speeds (in km/h) estimated by the three internet journey planners (see Section 2.2) for the journey from urban area *i* to urban area *j*

α, β, γ, ρ, ν, ξ and *ζ* are parameters determined during calibration

Table 2: Summarised Results From Calibration of Model (2)

	Value	t stat
Intercept	17.150	20.721
Road distance (ρ) parameter	-2.309	-15.589

Rail fare (v) parameter	-0.701	-3.475
Rail speed (ξ) parameter	2.050	9.279
Car speed (ζ) parameter	0.359	0.803
R_{adj}^2	0.970	

The inclusion of a bus speed variable to account for the effect of bus and coach competition was tested, but the resulting parameter was not of the expected sign, indicating that rail demand increases as bus speed increases. Model (2) was therefore retained as the preferred model. In a model of this form the elasticities are equivalent to the parameter values, and the rail fare elasticity is within the range of values currently used for demand forecasting (Association of Train Operating Companies, 2013). The formulation of the other parameters prevents direct comparison with previous research, with most other work using journey time rather than speed as an explanatory variable, but the values obtained from the model seem to be plausible.

In order to assess the extent to which there is potential for increasing rail use by improving the rail service (or to which rail use is suppressed by poor rail services), the predictions from Model (1) and Model (2) were compared to establish how much expected demand is affected by the quality of the rail service on offer. This was investigated by calculating the percentage difference between the predictions from the two models, with these differences illustrated in Figure 4. Such differences reflect the extent to which the relative quality of the rail service on a particular flow causes rail use to vary from the level expected given the characteristics and proximity of the urban areas linked by the flow. These differences are illustrated in Figure 4, which indicates that, as would have been intuitively expected, rail use is suppressed most on flows where the rail service offer is relatively poor. Table 3 gives details of the flows where the quality of the rail service means that rail demand is predicted to be less than 50% of what might otherwise be expected. The table shows that, in all cases, car journey times are significantly less than those by rail, a point backed up by Figure 5 which shows the ratio of rail generalised journey time to car journey time for the study flows.

Fig. 4 Percentage Differences Between Predicted Rail Use From Model (2) and Model (1)

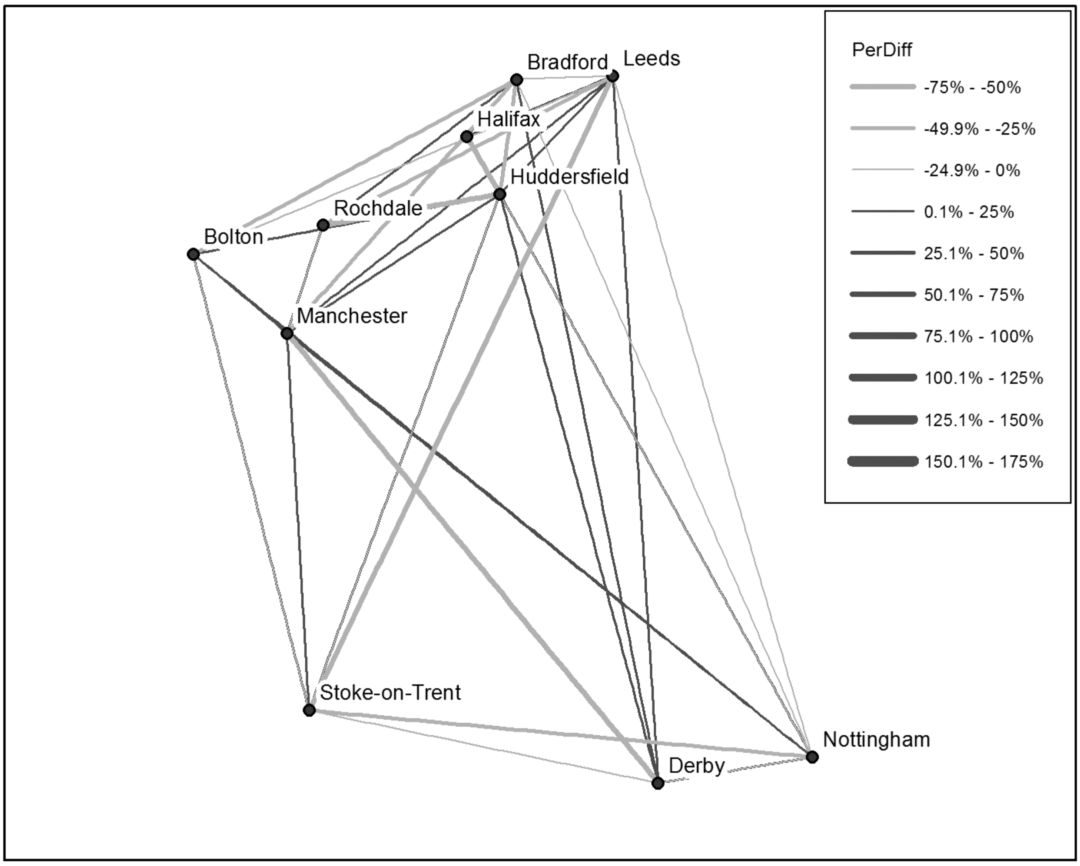
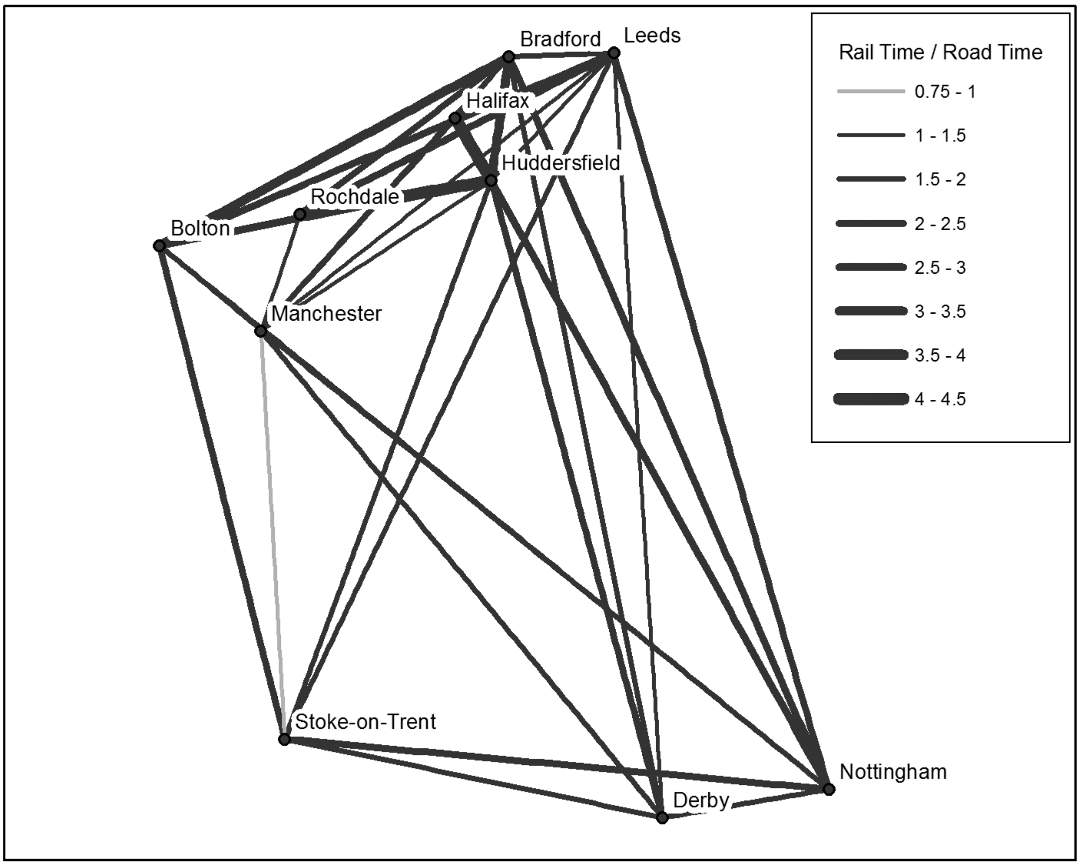


Table 3: Flows Where Relatively Low Rail Service Quality May Suppress Demand By 50%+

	Origin	Huddersfield	Halifax	Rochdale	Leeds	Manchester	Derby
	Destination	Halifax	Huddersfield	Huddersfield	Stoke	Derby	Manchester
Trips	Actual	8867	10312	1444	4614	16175	25238
	Model (1)	29199.3	34363.2	2868.4	9196.9	29301.7	66507.5
	Model (2)	9106.6	11246.4	1160.4	3916.6	12961.4	30062.1
	% Diff.	-68.9%	-67.2%	-59.6%	-57.4%	-55.7%	-54.8%
Distance (km)		11.4	11.4	27.6	114.3	91	91
Rail time (mins)		24	23	93	129	138	145
Rail GJT (mins)		58.6	62.9	121.5	159.2	149.7	159.9
Car time (mins)		16.3	15.7	28.3	98.3	105.7	105.7

Fig. 5 Ratio Of Rail Journey Times To Car Journey Times



The models can now be used to estimate the change in rail service quality that would be required to bring rail demand on a particular flow up to the mean level based on the origin and destination characteristics and proximity (ie to bring the Model (2) prediction up to the level of the Model (1) prediction). For example, for the flow from Rochdale to Huddersfield, it would be necessary to increase the number of trips by 1,708 to bring rail use up to the mean level, and the GJT required to give such an increase can be calculated using Equation (3). A full description of this application of the methodology is given in Section 4.3.

$$GJT_{req} = K_{ij} / \left(\sqrt[2.05]{\frac{T_{ij1}}{\left(\frac{T_{ij2}}{RsGJT_{ijo}^{2.05}} \right)}} \right) \quad (3)$$

Where:

GJT_{req} is the GJT required as an input for Model (2) to produce the same demand forecast as Model (1), with all other inputs held constant to current levels

T_{ij1} is the number of rail trips forecast from urban area i to urban area j by Model (1)

T_{ij2} is the number of rail trips forecast from urban area i to urban area j by Model (2)

$RsGJT_{ijo}$ is the observed mean rail generalised journey time per kilometre from urban area *i* to urban area *j*

The results from this demand-side analysis were combined with those from the capacity analysis described in Section 3 to identify which of the flows identified as underperforming in terms of rail demand (and therefore with latent demand for rail travel if services improved) have existing spare capacity to allow the rail service offer to be enhanced. This synthesis is described in Section 4.

3 CAPACITY ANALYSIS

3.1 Capacity and capacity utilisation

Railway capacity is an increasingly scarce and valuable commodity, and also a somewhat elusive measure, in that it is difficult, if not impossible, to define a specific capacity value for any given track, route or network (UIC, 2004; RSSB, 2010). Instead, the achievable capacity depends upon how infrastructure is used, and depends not only upon the characteristics of the track and signalling systems but also upon the performance characteristics and mix of trains using the route, the timetable, and the required punctuality level. Capacity Utilisation assessments are used to help infrastructure managers make best use of the capacity available while maintaining service quality and reliability. Where capacity is scarce, the objective will typically be to maximise capacity utilisation while still maintaining a stable, reliable quality of service. In order to achieve this, it is necessary to have measures of both provided capacity and capacity utilisation for all parts of the route or network in question, i.e. at nodes (junctions and stations) as well as links (sections of plain line between nodes). The UIC 406 (UIC, 2004) and Capacity Utilisation Index (CUI) (Gibson et al., 2002) approaches provide well-defined capacity utilisation measures based on timetable compression techniques for links, but not for nodes. This is a significant shortcoming, which this research addressed by extending the CUI approach to cover both junction and station nodes.

3.2 Extending the CUI approach from links to nodes

Considering junctions first, these vary from the simplest turnout to situations where multiple, multi-track routes converge, cross and diverge. Network Rail's Timetable Planning Rules (Network Rail, 2013) specify minimum junction margins for all junctions, giving the minimum time period permissible between train movements through the junction. These are equivalent to minimum headway values for links, meaning that base capacity utilisation values for a single junction can be calculated by 'compressing' all train movements so that the interval between successive movements is reduced to the minimum junction margin, ignoring any effects on the adjacent line sections. In reality, and as cautioned in the Timetable Planning Rules, operating multiple successive trains through the junction at minimum margins is likely to affect service reliability and quality. Suitable additional margins for junction movements should be defined using historic performance data to assess stochastic delays incurred at junctions where successive services operate at minimum margins. This approach can provide the basic 'building block' for assessment of complex junctions which are treated as a collection of simple nodes (and links), for each of which an individual capacity utilisation value could be calculated, and then combined to give an average capacity utilisation value for the overall junction. Colour coding could be used to illustrate those areas of the junction that are most heavily and lightly utilised, and between which a transfer of traffic might be considered in order to optimise utilisation.

Moving on to stations, the fundamental distinction between a station and a section of plain line is that some trains are timetabled to stop at a station, introducing time segments (dwell times) where the trains are scheduled to be stationary. Simple, through stations with one or two platforms are thus essentially similar to sections of plain line, while as station layout complexity increases, and individual approach tracks serve multiple platforms, or single platforms are connected to multiple departure tracks, stations come to resemble junctions. A further element of complexity is introduced when service patterns mean that trains change direction within a station, whether by necessity at a bay platform or by choice at a through platform. Minimum platform reoccupation times are the minimum headway equivalent here, with these being used together with station dwell times to ‘compress’ train movements giving a base capacity utilisation value. Further refinements are necessary to allow the assessment of more complex station layouts, for example by incorporating requirements such as the minimum time needed to clear a route out of a terminus platform after an adjacent platform has been occupied by a train using a shared approach track. However, as with junctions it is possible to treat complex station layouts, comprising multiple platforms and nodes, as a collection of simple platforms/stopping locations, nodes and links, for each of which an individual capacity utilisation value could be calculated. This approach therefore fills a gap in the range of capacity assessment tools available, and is similar in principle to the analytical approach to the capacity assessment of interlockings, as described by Pachl (2009), whereby an interlocking layout is “divided into small layout elements, which may be regarded as single channel systems.” A significant potential advantage of the approach is the avoidance of the need for detailed and time-consuming micro-simulation of a wide range of preliminary infrastructure and/or timetable options in the early stages of the development of an enhancement project.

4 SYNTHESIS AND APPLICATION

4.1 Case study route

The service provision and capacity assessment methodologies were brought together to produce a proposed approach for identifying the potential for the use of spare network capacity to provide an enhanced rail service, and this is again illustrated using a UK example. The route between Huddersfield and Rochdale in the north of England was chosen from the set of urban area pairs identified in Section 2 where current levels of rail use are lower than might be expected given the characteristics of the urban areas and their spatial separation. While this route had a lower degree of suppression than some others, the rail network connecting Huddersfield and Rochdale is relatively simple (see Figure 5), and its choice as an initial case study avoided the need to model highly complex junction layouts. The rail route between Huddersfield and Rochdale is however circuitous, reflecting the local terrain, with the rail distance approximately twice the 'crow fly' distance, and there are currently no through services between the towns.

4.2 Capacity analysis

Existing timetable data for the route were used to create a timetable graph in MS Excel, using an automatic timetable import and train graphing tool developed by Armstrong et al. (2009). Freight traffic on this route is limited and was excluded from the analysis, based on the assumption that any daytime freight paths that exist could be shifted and/or amended to accommodate additional passenger services. The objective of the exercise was to provide an additional, approximately hourly, service in each direction between Huddersfield and Rochdale. The options available for achieving this were constrained by the further aim of providing the service using just two additional rolling stock units to minimise costs, with an inbound service in one direction forming the following, outbound service. Since there are no existing direct services between Huddersfield and Rochdale, the timings of different trains over the various components of the overall route were combined to produce new, 'composite' trains in each direction, with individual sectional running times and speeds derived from existing services, and existing minimum headways, junction margins, and platform re-occupation and turnaround times (Network Rail, 2013). Combining the available train paths with minimum turnaround times at each end of the route, while maintaining minimum headway values along the route at all times, proved to be a considerable challenge, but a near solution was found giving broadly consistent hourly departure times throughout the day.

4.3 Demand impacts

The number of rail trips predicted by Model (2) (which takes account of rail service quality) from Rochdale to Huddersfield was 1160 (compared to an observed total of 1444), whereas Model (1) predicted 2868 trips based only on the characteristics of the origin and destination and their relative proximity. Equation (3) indicated that a reduction in GJT from 121.5 to 78.4 minutes was needed to increase demand to Model (1) levels. Calculating the GJT value which would result from the introduction of the service described in Section 4.2 is not entirely straightforward, as this service would operate alongside the existing half-hourly service requiring interchange. There are two possible options for estimating the new GJT value, with the first being to assume that the direct service replaces the slower interchange-based option. In this case the service interval penalty remains the same at 23 minutes (because the overall service frequency is unchanged), but the average journey time is reduced to 61.75 minutes, and the average interchange penalty to 8 minutes, giving a total GJT of 92.75 minutes. Substituting this value back into Model (2) gives a prediction of 2013 trips, still some way short of the Model (1) prediction. The second option is to treat the hourly direct service as the only service on offer, which means that the service interval penalty increases to 31 minutes (because the overall service frequency is reduced), the average journey time is reduced to 54.5 minutes, and the interchange penalty disappears, giving a total GJT of 85.5 minutes, superior to that obtained when the whole service on offer is considered but still 7 minutes higher than the target. A passenger survey would be required to establish which GJT value is more realistic, as this depends on whether passengers would perceive the direct service as being the only rail travel option, or whether they would still view the additional options requiring interchange as an integral part of the service option. This reflects a more general problem with calculating GJT values for routes where there is a high degree of heterogeneity in the rail service offer. Substituting the lower GJT value into Model (2) gives a prediction of 2390 trips, which is still almost 500 trips (or 17%) lower than that which would be expected given the urban area characteristics and their proximity. This probably results from the relatively indirect nature of the rail route between Rochdale and Huddersfield, which means that the rail distance is much longer than the road distance, and therefore that rail speeds have to be much higher than road speeds if rail is to be competitive for passengers who have a genuine choice of modes. It should also be noted that the number of additional trips per year generated by this service between Rochdale and Huddersfield would not in themselves be sufficient to justify the operation of an additional hourly direct service, and it would therefore be necessary to demonstrate that the service would generate large numbers of additional passengers on other flows in order to make a case for it to be introduced. While regional rail services are commonly subsidised in many countries, it will still always be necessary to balance the revenue and wider social benefits associated with generating new rail traffic against the costs of operating additional services, and the issue of cost modelling is considered in the following section.

5 COMPARISON OF COSTS AND REVENUES

Research which examines the demand and cost implications of new services, thus enabling an overall cost benefit assessment to be made, is particularly relevant given the current fiscal constraints in many countries, with transport policy emphasising the need for railways to reduce costs in order to earn their ‘licence to grow’. It would therefore be possible to identify services that would not be viable at present cost levels but would become profitable and/or socially beneficial if costs are reduced, for example by meeting the targets identified in the McNulty review of rail industry cost in the UK (McNulty, 2011), or through a more targeted approach to infrastructure maintenance and renewals on lightly used parts of the network.

Accurately modelling the costs of operating additional services is far from straightforward. However, there is some evidence from the previous literature on the impact of passenger rail franchising on costs and efficiency (Smith et al., 2010). This includes details of the extent to which costs vary in response to their key determinants and to which there are scale and

density economies in passenger operations in Britain (Wheat & Smith, 2010). It is therefore possible to make an approximate estimate of the costs of introducing new services based either on average costs or on marginal costs. In this case the latter are most appropriate, because all rail routes used by the additional services are already used by existing services, and a large proportion of the infrastructure costs will therefore be fixed. The marginal costs of operating additional services can be estimated using a cost elasticity with respect to train density to estimate the marginal costs of operating the additional services. Previous research in the British TOC sector by Smith and Wheat (2009), Wheat and Smith (2010), and Smith and Wheat (2011) suggests that an elasticity of around 0.8 is appropriate for TOC costs (in other words if the volume of services increases by 10%, total operating costs would increase by 8%). Based on data on variable access charges and total Network Rail costs in 2009/10, we estimated variable costs for the rail infrastructure to be around 6% of total costs (based on Network Rail Regulatory Accounts and Network Rail Annual Returns).
Introducing the additional train services proposed here would increase the Northern TOC's train km from 45.7 million to 46.18 million (a growth of 1.1%) and therefore, assuming the relationships described above are valid, TOC costs would be expected to increase by £3.45 million (from a base of £410 million) and Network Rail costs by £0.31 million (from a base of £515 million), giving a total additional cost of £3.75 million.

While further modelling work would be required to assess the full demand impact of service alterations, given that passengers on a large number of flows might make use of this service, it is possible to use the demand calculations from Section 4.3 to make an approximate estimate of the additional revenue generated. Fare data and the proportion of passengers currently using different ticket types were obtained from the TOAD dataset, and the same proportions were applied to the predicted new trip total to obtain a revenue estimate. For the Huddersfield-Rochdale service, the additional traffic generated on the Huddersfield-Rochdale flow would only translate into £10,500 of additional revenue, which would therefore not even cover 1% of the costs of operation. It seems very unlikely that there are over 100 additional and equivalent flows which would make use of this route and enable the service to break even. However, it should be noted that the relevant TOC (Northern Rail) received subsidies of £343 million in 2010-11 (Northern Rail Statutory Accounts[1]; this includes central government grants as well as those from Passenger Transport Executives (PTEs)), and the notional allocation of Network Rail direct grants to Northern Rail was an additional £339m in that year. This means that some existing services may require an even greater level of subsidy per passenger kilometre, and a sensible next step would therefore be to calculate the subsidy per passenger kilometre for existing services to establish whether reallocating some services from existing routes to new routes could increase the number of passenger kilometres accommodated within the existing budget constraint. Such a reallocation of resources may in any case prove more feasible than the introduction of extra services given the current restricted availability of additional rolling stock in the UK. Slightly better results were also obtained for some of the other flows in the case study dataset, with, for example, an hourly direct service from Manchester to Derby expected to generate £353,436 in additional revenue, against marginal additional costs of £9.19 million, although the costs still far exceed the revenue generated.

Additional farebox revenue is however not the only benefit which would be derived from such additional services. One of the main motivations of this study was the need to reduce road congestion and transport carbon emissions. By generating mode shift new services would generate benefits in both these areas, along with the wider economic benefits associated with improved rail services (Graham, 2007), benefits accruing to existing users on the route as a result of reduced travel times, and savings from reduced levels of road noise, vibration and accidents. Whether or not these benefits are sufficient to make a positive case for service expansion would need to be assessed on a case by case basis using standard appraisal methods, such as WebTAG in the UK (Department for Transport, 2012a), but the methodology outlined in this paper can play an important role in identifying the most

[1] Note that some adjustments have been made to align financial years to a March year end.

promising routes for such assessment. Given the planned construction of new high speed rail routes in the UK (Department for Transport, 2010) and elsewhere, with an associated release of capacity on conventional lines, the importance of reassessing service patterns in a number of countries is only likely to increase in coming years.

6 CONCLUSIONS AND DISCUSSION

This study has successfully developed a methodology which allows the identification of inter-urban flows where relatively low rail service quality appears to be suppressing rail demand, and therefore where the provision of a faster or more frequent rail service might generate significant modal shift from road. The application of this methodology to a case study of ten urban areas has shown that on some of the flows between these urban areas rail demand would be expected, all else being equal, to be significantly higher than is currently the case given the characteristics and proximity of the urban areas in question. Extending the research by applying the methodology to a larger matrix of flows would significantly improve the robustness of the results, and the increased size of the case study dataset would, by increasing the model degrees of freedom, in all likelihood allow significant parameters to be obtained for a much larger number of variables. This would mean, for example, that the quality of the rail service offer could be much more precisely defined through the inclusion of variables such as the hours of service operation. The scale of the data collection exercise required to allow such extensive analysis would be considerable, but once the dataset had been produced it could be suitable for use in a wide range of research areas. Improved data on travel volumes by modes other than rail would also enable the methodologies to be significantly enhanced, and the authors would therefore urge the collation of such data by the relevant industry and government bodies in all countries where mode shift to rail is a transport policy goal.

Alongside this demand-side methodology, methods for assessing the capacity utilisation of junctions and stations have also been developed. These form the basis for the development of a more generalised and generic approach to the capacity utilisation assessment of railway network nodes of varying size and complexity, while avoiding the need for detailed and costly simulation work in the preliminary stages of project development. This can provide the means of identifying locations where and times when spare network capacity is available, and may be used to provide additional passenger and freight capacity without the need for investment in additional infrastructure.

Combination of the demand and capacity analysis allows the identification of underperforming inter-urban flows where capacity exists to allow the rail service offer to be enhanced. While, in some cases, significant infrastructure investment would be needed to allow an improved service to operate, in other cases 'spare' infrastructure capacity would allow the operation of an improved service with minimal investment, providing potential 'quick wins' in achieving mode shift. Even if infrastructure for improved services is available, the costs of service operation may still be considerable, but these costs can be traded off against the more general reduction in unit costs resulting from the economies of density which appear to apply to rail operations. There could also be potential for generating additional traffic by increasing service frequencies at intermediate stations, which is not considered in the methodology described here, and from providing a greater level of service differentiation (for example by separating long distance and local flows). It seems feasible to combine the demand side analysis with existing cost data and models, and a full appraisal of the potential for improvements on the flows identified as the best options by this study would be an essential prerequisite for initiating service enhancements. While this methodology has been illustrated using a UK case study, the methods used should be transferable elsewhere, and a similar analysis could (and arguably should) be conducted in other countries with an established rail network, particularly where new high speed routes are planned.

Acknowledgments

This work was funded by the Engineering and Physical Sciences Research Council as part of the Cross-Disciplinary Feasibility Account titled 'Factor 20 – Reducing CO_2 Emissions From Inland Transport by a Major Modal Shift to Rail', ref EP/H024743/1. Thanks are due to Adrian Hickford at TRG Southampton for work on data collection for the demand modelling, to Jake Cartmell and Samantha Evens at the UK Department for Transport for providing access to and assisting with the extraction of the TOAD data used in this study, to Richard Batley at ITS Leeds for investigating data availability, and to all those involved in the initial discussion of this research at the RRUK Feasibility Account Sandpit in February 2010.

REFERENCES

Armstrong, J., Hood, I. and Preston, J.M. (2009) *Automating the Production of Train Graphs and the Calculation of CUI Values*. Proceedings of the 3rd International Seminar on Railway Operations Modelling and Analysis: Zurich.

Arup & Oxera (2010) *Is the Data Capable of Meeting the Study Objectives? Data Capability Report*. Oxford: Oxera Consulting Ltd.

Association of Train Operating Companies (2013) *Passenger Demand Forecasting Handbook v5.1*. London: ATOC.

Blainey, S.P. & Preston, J.M. (2009) Modelling Local Rail Demand In South Wales. *J. of Trans. Plan. and Technol.*, 33(1):55-73.

Blainey, S.P. and Preston, J.M. (2013) A GIS-based appraisal framework for new local railway stations and services. *Trans. Policy*, 25:41-51.

Booz & Co. (2010). *Costs of Railway Outputs*, Report for Rail Value for Money Team, London: Booz & Co.

Brons M, Givoni M and Rietveld P (2009) 'Access to Railway Stations and its Potential in Increasing Rail Use', *Trans. Res. A* 43:136-149

Department for Transport (2010) *High Speed Rail*, DfT, London

Department for Transport (2011) *Rail Subsidy Per Passenger Mile*, DfT, London

Department for Transport (2012a) *TAG Unit 3.13.1 Guidance on Rail Appraisal*, DfT, London

Department for Transport (2012b) *Transport Statistics Great Britain: 2012*, DfT, London

Gibson, S., Cooper, G., Ball, B. (2002) Developments in Transport Policy: The Evolution of Capacity Charges on the UK Rail Network, *J. of Trans. Econ. and Policy*, 36(2):341-354

Graham, D.J. (2007) Agglomeration, Productivity and Transport Investment, *J. of Trans. Econ. and Policy* 41(3):317-343

Lane, C., DiCarlantonio, M. U., Usvyat, L. (2006) Sketch models to forecast commuter and light rail ridership: update to TCRP report 16. *Transp. Res. Rec.*1986:198-210

McNulty, R (ed.) (2011) *Realising the Potential of GB Rail: Final Independent Report of the Rail Value for Money Study (Detailed Report)*. London: Department for Transport / Office of Rail Regulation

Network Rail (2007) *North West Route Utilisation Strategy*. London: Network Rail.

Network Rail (2009) *Yorkshire and Humber Route Utilisation Strategy*. London: Network Rail.

Network Rail (2010) *East Midlands Route Utilisation Strategy*. London: Network Rail.

Network Rail (2013) *Timetable Planning Rules*. http://www.networkrail.co.uk/browse%20documents/Rules%20Of%20The%20Route/roprhome.pdf Accessed 28 August 2013.

Office of Rail Regulation (2011) *National Rail Trends 2010-11 Yearbook*. London: Office of Rail Regulation.

Pachl, J. (2009) *Railway Operation and Control*. 2nd ed. Mountlake Terrace: VTD Rail.

RSSB (2010) *Creative Continuity – Exploring mega-city suburban network segregation*. http://www.rssb.co.uk/sitecollectiondocuments/pdf/reports/research/T915_rpt_final.pdf Accessed 12 June 2013.

Smith, A.S.J. & Wheat, P.E. (2009) The Effect of Franchising on Cost Efficiency: Evidence From the Passenger Rail Sector in Britain. In *11th Conference on Competition and Ownership in Land Passenger Transport*. Delft University of Technology, 2009.

Smith, A.S.J., Wheat, P.E. & Nash, C.A. (2010) Exploring the Effects of Passenger Rail Franchising in Britain: Evidence From The First Two Rounds of Franchising. *Res. in Trans. Econ.*, 29:72-79.

Smith, A.S.J. & Wheat, P.E. (2011) Evaluating Alternative Policy Responses to Franchise Failure: Evidence From the Passenger Rail Sector in Britain. *J. of Trans. Econ. and Policy*.

Thøgersen J (2006) Understanding Repetitive Travel Mode Choices in a Stable Context: A Panel Study Approach, *Trans. Res. A* 40(8):621-638

UIC (2010) *Leaflet 406: Capacity*. Paris: UIC (International Union of Railways)

Wardman, M., Lythgoe, W.F. and Whelan, G. (2007) ‘Rail passenger demand forecasting: cross-sectional models revisited’, in Dennis, S.M. and Talley, W.K. (eds) *Research in Transportation Economics*, Elsevier, Amsterdam, ch. 5.

Wheat, P.E. & Smith, A.S.J. (2010) *Econometric Evidence on Train Operating Company Size*. Office of Rail Regulation.