

AN AUTOMATED ONLINE TOOL TO FORECAST DEMAND FOR NEW RAILWAY STATIONS AND ANALYSE POTENTIAL ABSTRACTION EFFECTS

Marcus A. Young
University of Southampton
Simon P. Blainey
University of Southampton
Tom Gowland
Science and Technology Facilities Council
Srikanth Nagella
Science and Technology Facilities Council

1. INTRODUCTION

This paper describes a national trip end model to forecast demand for new local railway stations in Great Britain, which has been implemented as an automated tool hosted on the Data and Analytics Facility for National Infrastructure (DAFNI). It represents continuing progress in the development of a national trip end model by researchers at the University of Southampton's Transportation Research Group (TRG) over recent years.

The number of rail passenger journeys is expected to increase by up to 40% by 2040 (Network Rail, 2018), and the UK Government has recently set out an ambition of 'reversing the historic contraction of the rail network' with an emphasis on new local connections and stations that support housing development or economic growth, or that address urban congestion (Department for Transport, 2017). There will, therefore, be a continuing need to assess proposals for new railway stations and lines. A crucial part of this evaluation process is to obtain accurate demand forecasts, as predicted station patronage is a key driver of the benefits that will determine whether a scheme is considered viable.

Trip rate or trip end models¹ are the most common type of model used to forecast demand for new railway stations in GB. They were used in 10 of the 18 schemes reviewed in a report commissioned by the Department for Transport (DfT) to investigate the accuracy of recent station demand forecasts (Steer Davies Gleave, 2010). In only three of these cases was observed demand within 20% of the forecast, with examples of substantial under- or over-prediction. For example, observed demand was 2.65 times higher than forecast for Glasshoughton and less than half that expected at Aylesbury Vale Parkway. More recently, the appraisal for the new Borders Railway line severely under-forecast demand at the three Borders stations, with actual demand up to eight times higher than forecast in the first 12 months of operation (Transport Scotland, 2017). The models are typically developed and applied on a local basis by the consultants commissioned to evaluate a proposed scheme, reflecting guidance from the

DfT (Department for Transport, 2011) and the rail industry (Association of Train Operating Companies, 2013), which both consider the appraisal of a new station to be a special case requiring a bespoke model. However, as these models are specific to a local context they may not have been rigorously evaluated and it can be difficult to know what confidence to have in the demand forecasts that they generate.

Previous work at TRG has successfully calibrated national trip end models suitable for general application in forecasting demand for new local rail stations in England and Wales (for example, see Blainey, 2010). However, a potential weakness of this previous work is the use of a simplistic catchment definition to calculate the population from which trips will be generated. This typically involves dividing the study area into zones and assigning each zone to its nearest station. For example, Blainey (2010) used census output areas, and the scheme appraisal for Mitcham Eastfields station, which opened in 2008, used census enumeration districts. This method produces discrete non-overlapping catchments which imply that station choice is a deterministic process (everybody beginning a trip from within a zone is assumed to board at the same station) and that stations do not compete for passengers. However, real station catchments are far more complex entities, and the failure to account for this could be a contributory factor in the poor performance of models used to appraise some recent schemes. If station catchments are not adequately defined, then inappropriate weight will be given to other explanatory variables in the model, such as service quality measures, as drivers of trip generation, rather than the catchment population. This can result in less robust models that have poor geographic transferability — an important consideration when the aim is to develop a nationally applicable model (Wardman & Whelan, 1999). To address this concern, a novel aspect of the trip end model described in this paper is the use of probabilistic station catchments defined using a station choice model.

The national trip end model is intended to be used during the appraisal of local rail schemes for new stations or new lines, or where a non-incremental change to services at an existing station or stations is proposed. It can be used during the early assessment of different options and as part of a sifting process, providing a forecast of trips for each option, potentially alongside an analysis of abstraction from existing stations. For later appraisal stages, the forecast can form a key input to the cost benefit analysis, enabling the change in train operator revenue to be estimated. Crucially, the ability to generate probabilistic station catchments that capture competition between stations allows an estimate of abstraction from existing stations to be made. The model reduces the need to design and implement local modelling solutions which will inevitably vary in approach, robustness and performance. Even if a bespoke local model was considered desirable, this national model could act as a useful sense-check of the demand forecast it generated.

2. THE MODEL

2.1 Station Choice Component

A multinomial logit (MNL) model was calibrated using observed choice data from origin-destination passenger surveys carried out in 2014 and 2015, which were obtained from the Welsh Government and Transport Scotland. Several novel techniques were developed to validate these datasets and maximise their usefulness. These included the estimation of trip origins from incomplete address information, and the automated identification of illogical trips. The cleaned dataset consisted of 14,422 choice situations. The choice set for each observation was specified as the ten nearest stations to the origin postcode. Predictor variables were derived from open transport data sources, and a processing framework based around OpenTripPlanner, R and PostgreSQL was implemented to manipulate the large amount of data in a reproducible manner (Young, 2016). An API wrapper was written to query OTP and parse the planner response (now available as the R package *otpr* on CRAN). The model selected for incorporation in the subsequent trip end model calibration had an adjusted R^2 of 0.71 and a predictive performance measure of 24.9% (where closer to zero is better²). The model performed considerably better than a comparator model (42.2%), where the nearest station to the trip origin was assumed be chosen (replicating the deterministic approach). The utility function of the model, for individual n at origin i choosing station k is as follows:

$$V_{nik} = \exp(\beta N_k + \gamma \sqrt{D_{ik}} + \delta U_k + \epsilon \ln F_k + \zeta C_k + \eta Ps_k + \theta T_k + \iota B_k),$$

where D is the access distance by road from origin i to station k ; F is the daily service frequency at station k ; Ps is the number of car parking spaces at station k ; N , U , C , T and B are dummy variables that take the value of 1 if station i is the nearest station, unstaffed, has CCTV, has a ticket machine, or has a bus interchange respectively, and zero otherwise; and β , γ , δ , ϵ , ζ , η , θ , and ι are the estimated parameters. A summary of the calibration result for this model is shown in Table 1.

Table 1 Summary of calibration result for the station choice model.

Variable	β	z	sig
Nearest by distance	0.691	18.4	***
Sqrt(distance)	-2.262	-56.3	***
Category F station (small, unstaffed)	-0.677	-16.0	***
Ln(service frequency)	1.199	34.6	***
CCTV	1.071	8.6	***
Car parking spaces (#)	0.001	16.5	***
Ticket machine	0.984	19.1	***
Bus interchange	0.758	13.6	***
Sample size	14422		
Initial log-likelihood3	-33025		
Final log-likelihood	-9651		
McFadden's adjusted R^2	0.71		

To assess the likely predictive accuracy of the model on new data, a 10-fold cross-validation repeated 10 times was carried out. The average predictive performance measure of all repeats was 28.6%, representing only a small reduction in model predictive performance of 3.7 percentage points compared to the in-sample assessment. There was also very low variance in the average predictive performance measure between repeats (maximum difference of 0.4), indicating a high level of model stability. An example visualisation of a probabilistic catchment generated using the model is shown in Figure 1. For more information about calibration of the station choice model see Young & Blainey (2018) and Young (2019).

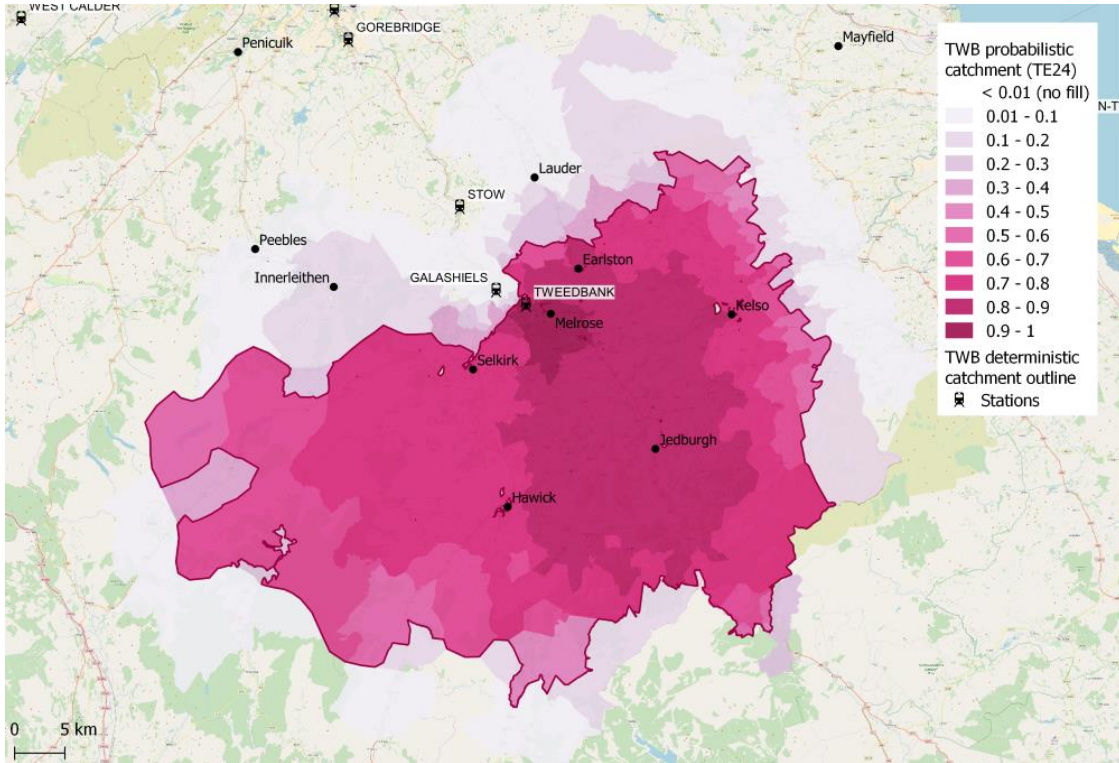


Figure 1 Deterministic and probabilistic catchments for Tweedbank station.

2.2 Trip End Model

The starting point for the trip end model calibration was a model developed during previous work by Blainey & Preston (2013) and Blainey (2017), where deterministic catchments were defined by assigning the population of census output areas to their nearest station. This model was modified to use the postcode as the zonal unit⁴, with the population of each postcode distributed to each station in its choice set, in proportion to the probability of each station being chosen. The form of the model is shown below:

$$\ln \hat{V}_i = \alpha + \beta (\ln \sum_z Pr_{zi} P_z w_{zi}) + \gamma \ln F_i + \delta \ln J_{it} + \epsilon \ln P_{Si} + \zeta T e_i + \eta E l_i + \theta B_i,$$

where \widehat{V}_i is the estimated annual passenger entries and exits for station i ; Pr_{zi} is the probability of someone located in postcode z choosing station i ; Z consists of all postcodes which have station i within their choice set; w_z is a two-stage distance decay function; F_i is weekday train frequency at station i ; J_{it} is the number of jobs within t minutes drive of station i , Ps_i is the number of parking spaces at station i , and Te_i , El_i and B_i are dummy variables that take the value of 1 if station i is a terminus station, served by electric trains or a travelcard boundary station respectively, and zero otherwise; and $\alpha, \beta, \gamma, \delta, \epsilon, \zeta, \eta$, and θ are the estimated parameters.

In line with the earlier work, the calibration dataset is defined as the majority of category E and F stations⁵ in Great Britain, and the dependent variable is the total number of station entries and exits in 2011/12 as reported by the Office of Rail and Road (2013). To generate the bracketed part of the equation, which represents the catchment definition, choice sets were constructed for every postcode in mainland GB (some 1.5 million), consisting of the ten nearest stations to each. The associated choice probabilities were then calculated using the station choice model discussed above. A summary of the calibration result for the trip end model selected for the automated implementation on DAFNI is shown in Table 2.

Table 2 Summary of calibration result for the trip end model.

<i>Variable</i>	β	z	<i>sig</i>
<i>Intercept</i>	3.67	38.5	***
<i>ln(population)</i>	0.37	20.14	***
<i>ln(daily train frequency)</i>	1.14	41.47	***
<i>ln(work population within 1 minutes' drive)</i>	0.05	7.75	***
<i>ln(car park spaces)</i>	0.13	14.14	***
<i>Electric services</i>	0.24	5.93	***
<i>Travelcard boundary</i>	0.30	3.29	**
<i>Terminus station</i>	0.78	9.37	***
<i>McFadden's adjusted R²</i>	0.85		
<i>Mean Squared Error (MSE)</i>	0.48		

The model was found to perform better, in terms of adjusted R² and AIC, than a comparative model with deterministic catchments (postcode population assigned to nearest station). Importantly, greater weight was given to the population variable and reduced weight was given to variables related to station services and characteristics. This indicates that the more realistic representation of the catchment enables differences in the number of trips to be better explained through the population variable. Consequently, the model should be more transferable and better suited for use as a national predictive model. To identify any potential systematic bias in the model at

regional level, the standardised residual for each station was plotted on a map of GB, as shown in Figure 2.

In this map the radius of each point is proportional to the size of the residual (note that the points for stations with very small residuals are not visible at this scale). Overall, the map shows that under-prediction and over-prediction occurs in all regions of the country and is present at a range of magnitudes. This suggests that the model performs similarly across the country, with no obvious regions where the model systematically under- or over- predicts station demand, and no regions where the standardised residuals appear systematically larger than in others. There is perhaps a tendency for under-prediction to dominate in the Greater London area. This would be expected given that there is no realistic alternative to public transport modes for travelling to/from central London and there is no variable that captures this additional generation effect in the model.

The likely predictive accuracy of the model on new data was assessed using a 10-fold cross-validation, repeated ten times. The average estimate of mean squared error (MSE) across all ten repeats was only marginally higher (0.003) than the internal MSE, suggesting that the model's predictive validity will hold when applied to new data. There was only a small variance in the cross-validation estimate across the repeats (maximum 0.002), indicating that the model has high stability.

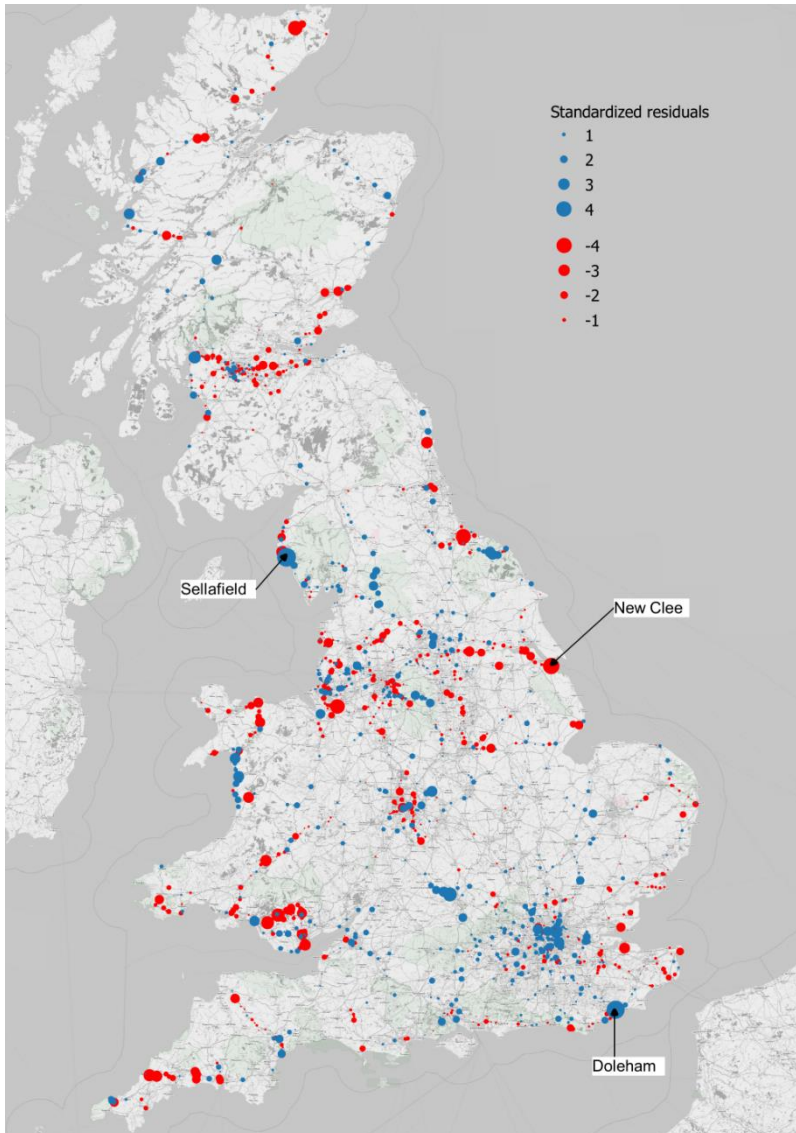


Figure 2 Standardised residuals for each station plotted on a map of GB. The radius of each point is proportional to the size of the residual with positive residuals (model under-prediction) shown in blue and negative residuals (model over-prediction) shown in red.

2.3 Evaluation

The model's predictive performance has been assessed for ten recently opened stations, including seven on a newly built railway line. For all but three stations, the model produced a more accurate forecast than the comparator model with deterministic catchments, highlighting the potential importance of using a trip end model that better represents real-life station catchments. The model also performed well when compared to the official forecasts produced during scheme appraisals, particularly for stations on the new Borders Railway line. The results of the evaluation are given in Table 3 and summarised in Figure 3.

Table 3 Demand forecasts for 10 recently opened stations and comparison with scheme forecasts and actual trips in 2017/18.

<i>Station</i>	<i>Year opened</i>	<i>Entries & exits 2017/18 (ORR)</i>	<i>Scheme forecast</i>	<i>% diff from 17/18</i>	<i>Forecast (simple catchment)</i>	<i>% diff from 17/18</i>	<i>Forecast (probabilistic catchment)</i>	<i>% diff from 17/18</i>	<i>Comment</i>
<i>Conon Bridge</i>	2013	15,100	36,000	138	24,453	62	25,091	66	Poor service performance suppressing demand.
<i>Energlyn & Churchill</i>	2013	101,362			73,015	-28	75,467	-26	
<i>Fishguard & Goodwick</i>	2012	20,136			14,345	-29	16,317	-19	
<i>Tweedbank</i>	2015	436,978	43,242	-91	806,146	84	520,157	19	Suppressed demand due to car park capacity.
<i>Galashiels</i>	2015	356,262	46,862	-44	200,381	-44	157,217	-56	Tourism/bus interchange generating extra trips.
<i>Stow</i>	2015	69,834	11,686	-83	96,263	38	77,841	11	
<i>Gorebridge</i>	2015	115,102	180,038	56	254,489	121	226,058	96	Potential bus competition.
<i>Newtongrange</i>	2015	157,016	105,836	-33	239,277	52	209,621	34	
<i>Eskbank</i>	2015	338,932	261,050	-23	312,784	-8	250,757	-26	
<i>Shawfair</i>	2015	31,588	123,720	292	106,627	238	65,467	107	Potential bus competition.

When comparing forecast demand with actual demand it should be noted that the trip end model has been calibrated using stations that, with a few exceptions, are well-established and have been open for many years. There is evidence that it can take several years (perhaps 5 or 6) for a new station to reach its potential as individuals adjust their behaviour over time (Preston & Dargay, 2005; Blainey, 2009). This should be borne in mind when comparing forecast demand with actual, especially in the initial years.

2.4 Abstraction Analysis

As a new station might abstract passengers from existing stations, the net additional demand realised could be substantially lower than the gross forecast might suggest. If the scheme appraisal process does not take this into account, it could result in a new station being built that fails to deliver the expected economic and societal benefits. To address this issue, a methodology has been developed to assess the potential extent of abstraction based on changes that occur to the probabilistic catchment(s) of the affected station(s). This has been incorporated as an optional analysis in the automated DAFNI model, and consists of the following main steps:

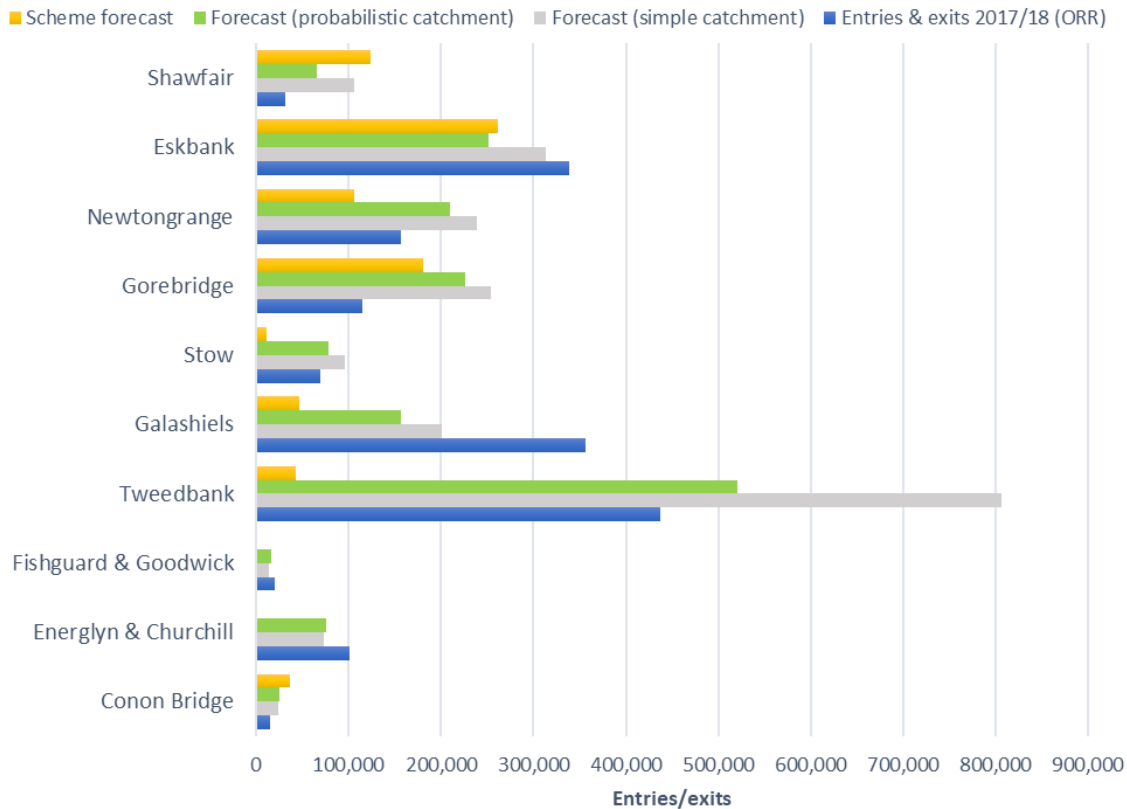


Figure 3 Summary of demand forecasts for 10 recently opened stations and comparison with scheme forecasts and actual trips in 2017/18

- For a station identified as being ‘at risk’ of abstraction, a ‘before’ choice set (derived from current stations only) and an ‘after’ choice set (derived from current stations *plus* the proposed new station) for each postcode⁶ are generated.
- The percentage change in the weighted population for the ‘at risk’ station, in the before and after situation, is calculated.
- Assuming an elasticity of one between weighted population and the number of entries/exits, the percentage change is applied to the latest annual entries/exists for the ‘at risk’ station, thus giving an estimate of the abstraction effect⁷.

This methodology has been used to assess abstraction resulting from several potential new stations in Wales, as part of consultancy work carried out for the Welsh Government. For example, the abstraction effect of a proposed new station known as ‘South Wrexham’ (located in Rhosymedre), on the existing stations at Ruabon and Chirk was analysed. This indicated that 40% of trips would be abstracted from Ruabon and Chirk. The effect of the new station on the probabilistic catchment for Ruabon station can be seen by comparing Figures 4 and 5, which show the catchment before and after the new station. While the proposed methodology has been successfully applied and appears promising, further work is needed to validate the approach.

For more information about the trip end model and its evaluation, see Young & Blainey (2018) and Young (2019).

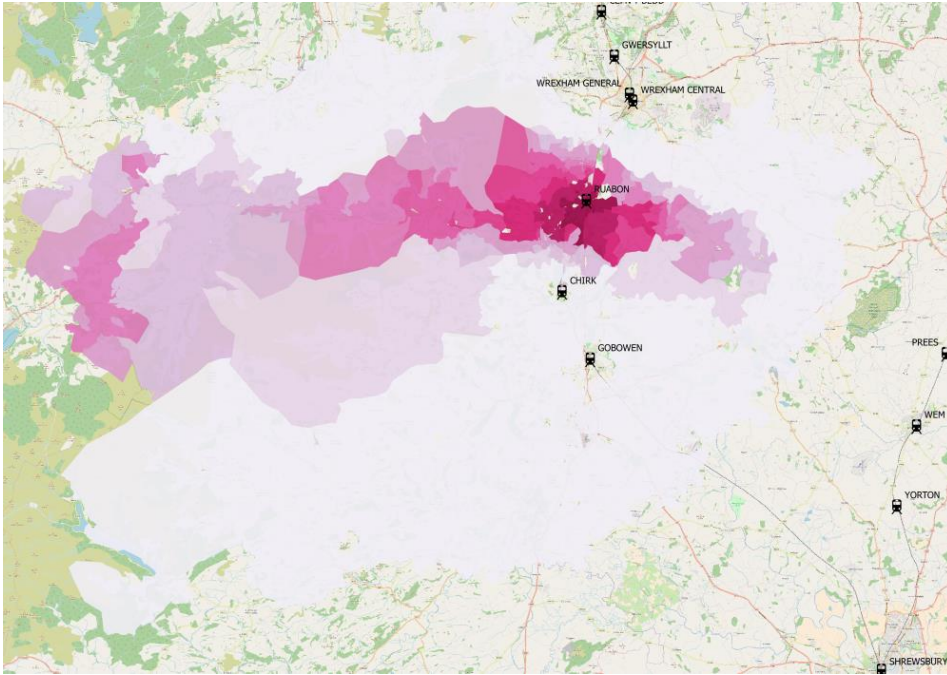


Figure 4 The existing probabilistic catchment for Ruabon station.

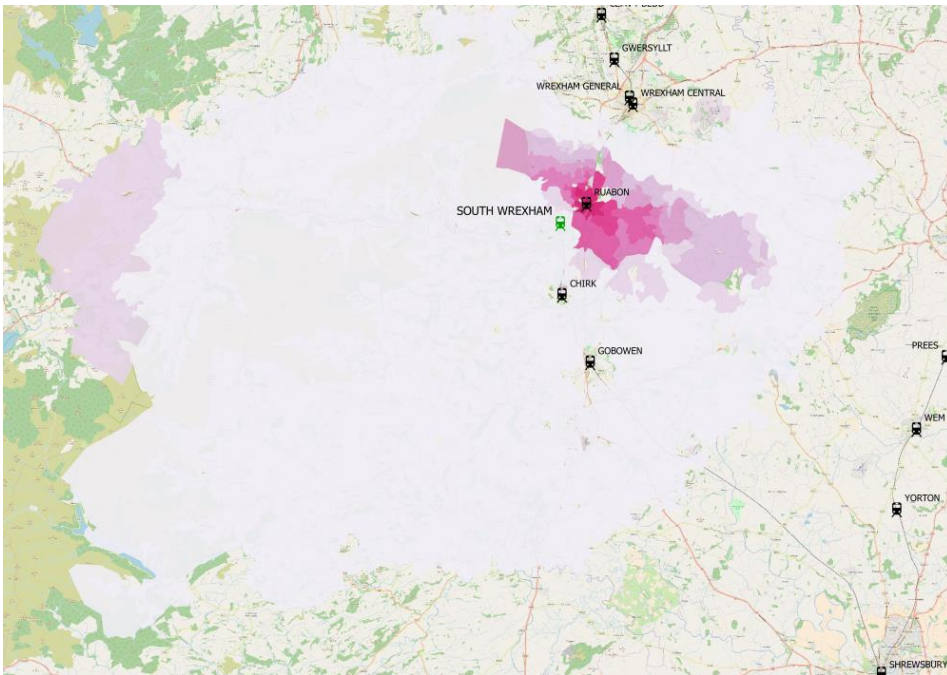


Figure 5 The probabilistic catchment for Ruabon station if South Wrexham station was opened.

3 IMPLEMENTATION ON DAFNI

3.1 About DAFNI

DAFNI is planned to become the UK's computational platform to support academic research into infrastructure systems. This research aims to deliver a national infrastructure system that is more efficient, reliable, resilient and affordable. DAFNI will provide data storage, data analytics, simulation, modelling and visualisation facilities and is funded by RCUK as part of the UK Collaboratorium for Research on Infrastructure and Cities (UKCRIC). Several pilot projects have been run alongside development of the core DAFNI platform, supported by software developers from the DAFNI pilot team. The projects have each implemented an existing infrastructure model in an environment that emulates the expected future DAFNI system, enabling system components to be validated and hardware to be stress-tested. The station demand forecasting model described in this paper was chosen as one of the pilot projects, with implementation funded via an EPSRC Impact Acceleration Account grant from the University of Southampton.

3.2 Automating the Model

Prior to beginning work on the pilot project, the model had only been run on a manual basis. Several stand-alone R scripts had been written to reproduce many of the necessary processing steps, but these still required significant manual intervention to make adjustments specific to each scenario and to run relevant code. In addition, the GIS analysis that was primarily used to create the station choice sets, was carried out manually using ArcGIS. The first stage of the project was, therefore, to create a fully automated model that was ready to be transferred to the DAFNI team for implementation. To avoid costly proprietary software licenses, it was a requirement that only open source software tools should be used. The automated model was primarily developed as an R package which was maintained on a private GitHub repository. Although not intended to be released to the wider R community, this enabled the model to be easily shared with the DAFNI team and imposed the rigour of established package conventions, such as organization of code and data, and providing comprehensive function documentation.

A significant challenge of the project was using open source tools to replicate the spatial analysis capabilities provided by ArcGIS, in particular the powerful origin-destination cost matrix tool of the Network Analyst extension, which had previously been used to generate the choice set of ten nearest stations for each postcode. This was ultimately achieved using pgRouting, which is an extension to a PostGIS/PostgreSQL spatial database that provides routing capabilities (see: <https://pgrouting.org>). However, as pgRouting does not support identifying the 'nearest x facilities' to an origin, a series of SQL wrapper functions were written to extend pgRouting's functionality and to optimise the routing performance on the large road network (Ordnance Survey Open Roads). This process has a very high computational cost and to reduce the time required for a model run it has been parallelised, taking advantage of the multiple processing cores available

for the container (virtualised operating system) running the model on DAFNI. This has delivered a step-change in performance, reducing the time to model a single station from around 30 minutes on a high-end workstation to less than five minutes. Time savings are substantially higher for larger and more complex model scenarios.

The model as hosted on DAFNI generates a demand forecast (predicted trips per year) for one or more proposed local railway stations. If required it can also produce an analysis of potential abstraction of journeys from existing stations, enabling the net impact of a new station on rail use to be estimated. Forecasts for multiple stations can be accommodated as part of the same job. These can be treated independently (*alternative* station locations are to be assessed) or concurrently (the proposed stations will *coexist*). The model provides a high degree of customisability, with the ability to adjust the service frequency of existing stations (for example if a proposal for a new station is accompanied by a planned increase in service level on a route), and to input exogenous data, such as new jobs, houses, and/or population.

3.3 Web-based User Interface

The DAFNI development team has provided a professional web interface that enables the user to interact with the model, delivers visualisation of outputs, and handles job management. It includes an interactive map that allows the user to select the location and road access point of a proposed station and to identify relevant postcodes or workplace zones to assign exogenous data to. It also has the facility to download or upload a model configuration file, providing further flexibility, such as the ability to make minor amendments to a prior job or automating generation of model inputs from other systems. The interface is intuitive to use, and no specialist knowledge or skills are necessary in order to prepare and submit a job to the model. A screenshot showing the form used to add stations to the model is shown in Figure 6.

DAFNI Pilot 3: Station demand forecasting model

ID: HEL51
 Name: Helston
 Region: South West
 Station NGR: Easting 166257, Northing 28051
 Access NGR: Easting 166257, Northing 28051
 Category: E
 Model abstraction for these stations: PNZ

Parking spaces: 25
 Frequency: 20
 Frequency group: grp1
 Ticket machine: Yes
 Bus interchange: No
 CCTV: Yes
 Terminal station: Yes
 Electric services: No
 Travelcard boundary: No

Buttons: ADD, RESET, EDIT (multiple)

For 'in isolation' mode only:
 If sensitivity analysis is required for service frequency or car parking spaces then simply create a new station entry with these fields amended as appropriate. The ID must be unique, but if you use the same name then the model results will be grouped and processing will be faster as unnecessary duplicate analysis will be avoided. If two or more entries have an identical name they can only differ in the values of the frequency, parking spaces and/or frequency group fields. Remember to do groups as appropriate for the additional entries.

ID ↑	Name	Region	Station NGR	Access NGR	Frequency	Frequency group	Parking spaces	Ticket Machine	Bus Interchange	CCTV	Terminal	Elect
HEL51	Helston	South West	166257 28051	166257 28051	20	grp1	25	Yes	No	Yes	Yes	No

Rows per page: 5

Figure 6 The DAFNI model web interface showing the proposed station input form.

4. CONCLUSIONS

Forecasting demand for new railway stations is considered by the rail industry to be a 'special case' requiring bespoke models to be developed and applied in a local context for the specific scheme being appraised. This is primarily achieved using trip rate/end models that have not always performed well. Given the background of growing passenger demand and increasing interest in opening new stations and lines, there will be an ongoing need to assess proposed schemes. The national trip end model that has been developed and implemented on DAFNI has the potential to remove the need for scheme proponents, such as local and regional government or transport authorities, to commission costly bespoke studies. In cases where it was still considered prudent to apply local models, this national model could be used as a sense-check tool. For example, if demand forecasts produced by the local and national models differed by an order of magnitude, it would be a clear warning that the local models may not be reliable. Given that the level of station usage is a key driver of the benefit-cost ratio upon which investment decisions are made, identifying a potential problem with the demand forecast at an early stage of a project would be hugely beneficial. Implementation of the model on DAFNI makes this powerful tool available to transport planning practitioners and other stakeholders for the first time. The data storage and transformation capabilities of DAFNI ensure that the model data inputs are always available and up-to-date, freeing practitioners from onerous collection and processing tasks. The tool has the potential to transform the assessment of new station schemes, enabling the rapid review of options for individual stations or new lines.

NOTES

1. Trip rate models assume the number of trips to be some function of the population in the area surrounding a station (its catchment), while trip end models include additional variables relating to station services, facilities or the locality.
2. Predictive performance was measured by comparing the sum of predicted probabilities for each station with the number of times that station was actually chosen. To assess the overall performance of a model, the absolute difference between the two figures was summed for all stations and expressed as a percentage of the total number of choice situations in the model. A 'predictive performance difference' of zero percent would therefore indicate no deviation between observed and predicted choice. There is no theoretical upper limit to the measure.
3. Initial log-likelihood assumes there is an equal probability of each alternative in a choice set being chosen.

4. The zonal system is therefore of much higher spatial resolution. There are some 230,000 census output areas in GB, but around 1.5 million postcodes.
5. Stations were divided into six categories (A – F) when the GB rail industry was privatised in 1996. Category A stations are national hubs, Category B are national interchanges, Category C are important feeder stations, Category D are medium staffed stations, and Category E and F are small stations, staffed and unstaffed respectively (Green & Hall, 2009).
6. Only postcodes that are within 60 minutes' drive time of a station are considered as potential candidate postcodes when calculating the probability weighted population to assign to a station. This is based on empirical evidence. This applies to both the station demand forecast and the abstraction analysis. Note: this does not mean that all postcodes within 60 minutes of a station are included. Only if the station of interest appears in a postcode's choice set (of ten nearest stations) will that postcode be considered.
7. This is based on evidence in the PDFH relating to the external environment and forecasting framework assumptions that the population elasticity is equal to one for the number of trips originating in a zone (Association of Train Operating Companies, 2013, Chapter C1). Although PDFH elasticities are intended to be applied to flows, the unitary elasticity assumes that only origin population can influence growth in rail demand. There is a lack of evidence on the appropriate elasticity to use if both origin and destination population changes are considered.

ACKNOWLEDGEMENTS

The support of the following data providers is gratefully acknowledged. The Welsh Government and Transport Scotland for providing passenger survey data; Code.Point and Code.Point Polygons © Crown Copyright and Database Right. Ordnance Survey (Digimap Licence). This work uses public sector information licensed under the Open Government Licence v3.0. Basemap data © OpenStreetMap contributors.

FUNDING

This work was supported by the EPSRC under DTG Grant EP/M50662X/1 and through the University of Southampton's Impact Acceleration Account.

BIBLIOGRAPHY

Association of Train Operating Companies. (2013). *Passenger demand forecasting handbook v5.1*.

- Blainey, S. (2009). *Forecasting the use of new local railway stations and services using GIS*, PhD Thesis, University of Southampton.
- Blainey, S. (2010). Trip end models of local rail demand in England and Wales. *Journal of Transport Geography*, **18(1)**, 153–165.
- Blainey, S. (2017). *A new station demand forecasting model for Wales*. (Unpublished working paper).
- Blainey, S., & Preston, J. (2013). A GIS-based appraisal framework for new local railway stations and services. *Transport Policy*, **25**, 41–51.
- Department for Transport. (2011). *Guidance note on passenger demand forecasting for third party funded local rail schemes*.
- Department for Transport. (2017). *Connecting people: a strategic vision for rail*.
- Green, C., & Hall, P. (2009). *Better rail stations* (An independent review presented to Lord Adonis, Secretary of State for Transport).
- Network Rail. (2018). *A better railway for a better Britain - strategic business plan 2019 - 2024*.
- Office of Rail and Road. (2013, May). *Estimates of station usage - 2011-12 report and data* [Excel spreadsheet]. Webpage. Retrieved from <http://orr.gov.uk/statistics/published-stats/station-usage-estimates>.
- Preston, J., & Dargay, J. (2005, October). *The dynamics of rail demand*. Paper presented at Third Conference on Railroad Industry Structure, Competition and Investment.
- Steer Davies Gleave. (2010). *Station usage and demand forecasts for newly opened railway lines and stations* (Final Report prepared for Department for Transport).
- Transport Scotland. (2017). *Borders Railway year 1 evaluation*. Retrieved 25 August 2017, from <https://www.transport.gov.scot/media/39335/sct04173824741.pdf>.
- Wardman, M., & Whelan, G. (1999). *Using geographical information systems to improve rail demand models*. (Final Report to Engineering and Physical Sciences Research Council).
- Young, M. (2016, March). *An automated framework to derive model variables from open transport data using R, PostgreSQL and OpenTripPlanner*. Paper presented at 24th GIS Research UK Conference.
- Young, M. (2019). *Modelling railway station choice: can probabilistic catchments improve demand forecasts for new stations?* PhD Thesis, University of Southampton.

Young, M., & Blainey, S. (2018). Development of railway station choice models to improve the representation of station catchments in rail demand models. *Transportation Planning and Technology*, **41(1)**, 80–103.