An automated online tool to forecast demand for new railway stations

Marcus Young
Simon Blainey
Transportation Research Group
University of Southampton
10 July 2019
What does the tool do?

Forecasts demand for new local railway stations

- Individual station(s)
- New line
- Abstraction analysis
- Catchment visualisation
Theoretical basis - a national trip end model incorporating probability-based catchments

\[
\ln \hat{V}_i = \alpha + \beta \left( \ln \sum_{z} Z Pr_{zi} P_{z} w_{zi} \right) + \gamma \ln F_i + \delta \ln J_{it} + \epsilon \ln Pk_i + \zeta Te_i + \eta El_i + \theta B_i
\]

- **V** - annual trips
- **Pr** - probability of station being chosen
- **P** - population
- **w** - decay function
- **Z** - postcodes with station \( i \) in choice set
- **z** - postcode

- **F** - service frequency
- **J** - jobs (within approx. 0.5 mile)
- **Pk** - parking spaces
- **B** - travelcard boundary (y/n)
- **Te** - terminus station (y/n)
- **El** - served by electric trains (y/n)
A station choice model calibrated using ~15,000 observations from on-train passenger surveys

\[ Pr_{nik} = \frac{\exp(\beta N_k + \gamma \sqrt{D_{ik}} + \delta U_k + \epsilon \ln F_k + \zeta C_k + \eta P_{Sk} + \theta T_k + \iota B_k)}{\sum_{k=1}^{K} \exp(\beta N_k + \gamma \sqrt{D_{ik}} + \delta U_k + \epsilon \ln F_k + \zeta C_k + \eta P_{Sk} + \theta T_k + \iota B_k)} \]

\( K = 10 \) nearest stations to each postcode

Pseudo \( R^2 = 0.71 \)
The trip end model was calibrated using Category E and F stations in mainland GB

\[ R^2 = 0.85 \]

\[ R^2 = 0.84 \]
Standardised residuals show that the model performs fairly consistently across the country
The model generally produces better forecasts than the scheme appraisal for 10 recently opened stations.
Abstraction analysis is based on expected change in an existing station’s probabilistic catchment

\[ \text{Trips}_{\text{old}} \times \left( \frac{\text{Pop}_{\text{w.prob.dist new}}}{\text{Pop}_{\text{w.prob.dist old}}} \right)^{1} = \text{Trips}_{\text{new}} \]
The tool has been implemented on the Data and Analytics Facility for National Infrastructure (DAFNI)

One of four initial pilot projects
The tool is built using open source tools and data

Identify all postcodes within 60 minutes of proposed station

Find the 10 nearest stations to each postcode

Calculate probabilities using station choice model

Run trip end model to generate forecast

Run abstraction analysis (if needed)
Generating choice sets is computationally intensive

- OS Open Roads
- Prepared for pgRouting
- Virtual nodes injected
- Station service areas created

During model run:
- Stations data
- Postcode centroids
- Virtual nodes injected for routing
- Use bounding box for performance
- Find first service area where postcode intersects >= 10 stations
- Get distance from postcode to each station

Service areas for Plymouth station

<table>
<thead>
<tr>
<th>Name</th>
<th>CRICODE</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAIGNTON</td>
<td>PGN</td>
<td>659</td>
</tr>
<tr>
<td>TORQUAY</td>
<td>TQV</td>
<td>3846</td>
</tr>
<tr>
<td>TORRE</td>
<td>TRR</td>
<td>5172</td>
</tr>
<tr>
<td>TOTNES</td>
<td>TOT</td>
<td>9804</td>
</tr>
<tr>
<td>NEWTON ABBOT</td>
<td>NTA</td>
<td>12585</td>
</tr>
<tr>
<td>TEIGNMOUTH</td>
<td>TGM</td>
<td>16487</td>
</tr>
<tr>
<td>DAWLISH</td>
<td>DWL</td>
<td>21040</td>
</tr>
<tr>
<td>DAWLISH WARREN</td>
<td>DWV</td>
<td>23888</td>
</tr>
<tr>
<td>IVYBRIDGE</td>
<td>IVY</td>
<td>27024</td>
</tr>
<tr>
<td>STARCROSS</td>
<td>SCS</td>
<td>27545</td>
</tr>
</tbody>
</table>

Query time: 0.215s  Record 1 of 10
DAFNI provides step-change in run times

Parallel processing employed in R

Each job runs in a separate container (virtualised OS)

Up to 30 cores per container

< 5 minutes for a single station forecast
The tool has a user friendly web interface

Includes input verification
Advanced front-end functionality is provided

- Interactive map
- Job submission
- Job management
- Logging
Demonstration – configuring and running a job

About the model

This DAFNI-hosted service 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 underlying model is based on research by the University of Southampton’s Transportation Research Group. At its core is a trip end model which has been calibrated on the smaller stations in Great Britain. In such a model the number of trips is a function of the population in a station’s catchment and a range of other variables, such as service frequency and number of jobs nearby. A novel aspect of this model is that probability-based catchments are defined using a station choice model. Rather than assuming everyone will use their nearest station, this provides a more realistic representation of behaviour and allows competition to occur between stations.

Generating a demand forecast involves a series of data processing and spatial analysis steps with a high computational requirement. As the model has been coded to take advantage of parallelisation, DAFNI provides an ideal environment, enabling it to run across multiple processor cores. This has delivered a step-change in performance, reducing the time to model a single station from around 60 minutes on a high-end workstation to some ___ minutes. Time savings will be substantially higher for larger and more complex model scenarios.

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. DAFNI has enabled what could otherwise have become a siloed model to be rapidly made accessible to other researchers and transport practitioners, thus maximising knowledge exchange and research impact.
In conclusion, the tool enables rapid review of scheme options for individual stations or new lines.

- Robust and transferable
- Alternative to costly bespoke models
- Useful as a sense-check of local models
- Incorporates abstraction analysis

Any Questions?

m.a.young@soton.ac.uk