UNIVERSITY OF Southampton

Developing railway station choice models to improve rail industry demand models.

Marcus Young
PhD student, Transportation Research Group
5 January 2017

UNIVERSITY OF Southampton

Outline

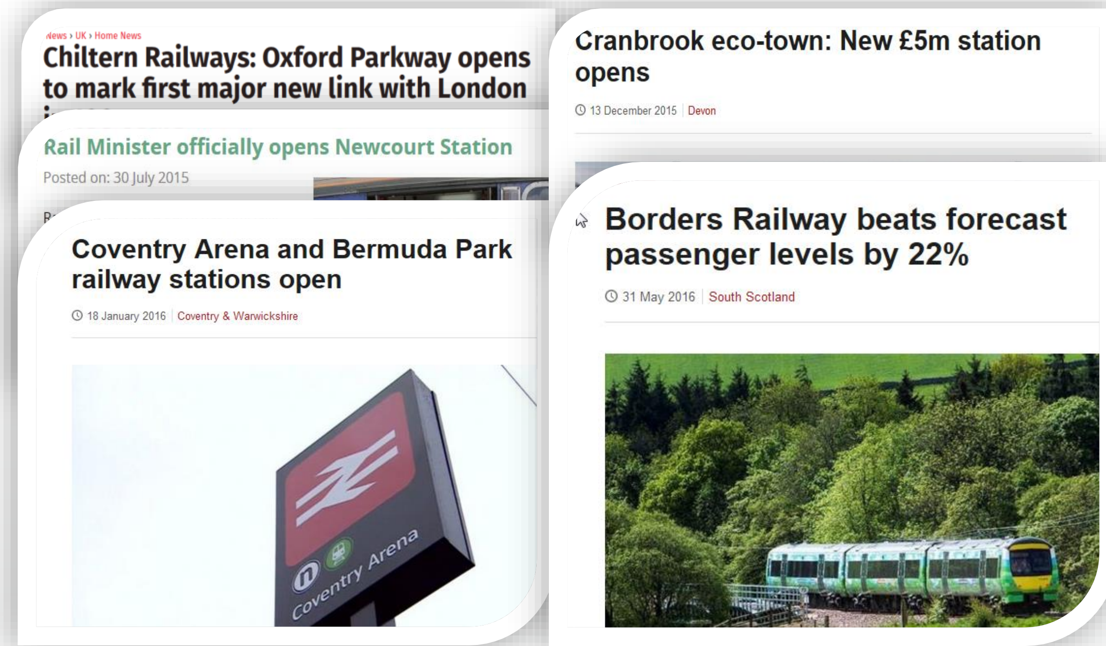
- Research background
- Data considerations
- Model results
- Model appraisal
- Conclusions & future work

Research background

A rail renaissance

UNIVERSITY OF Southampton

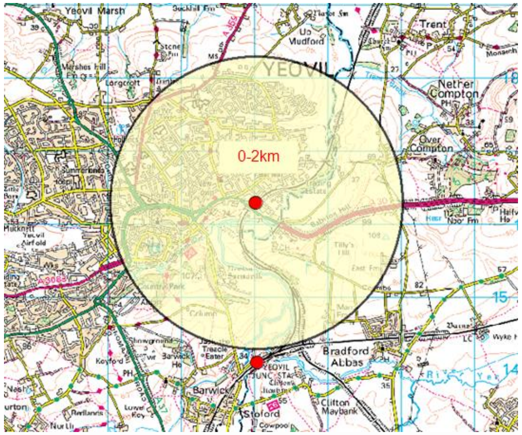
New stations

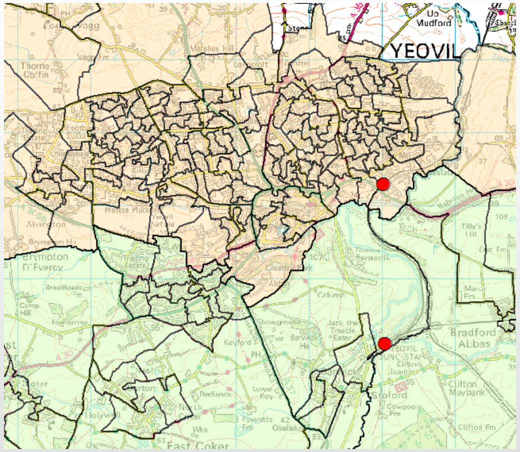


- Increasing interest in using rail to meet transport needs or drive economic growth
- Need accurate demand forecasts

Demand models – defining catchments

- Trip end and flow models
- Must define a catchment first:
  - circular (buffer) around station
  - nearest station – zone based
- Choice of station is deterministic
- Catchments are discrete, none overlapping



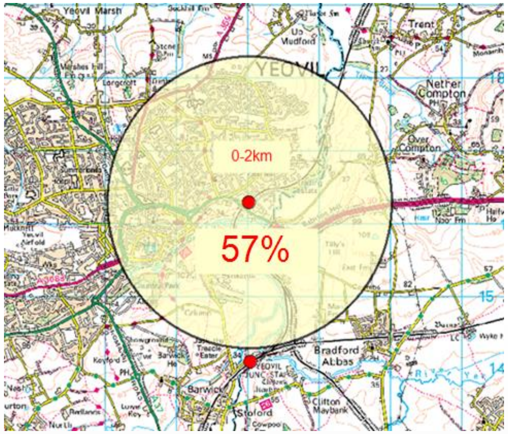


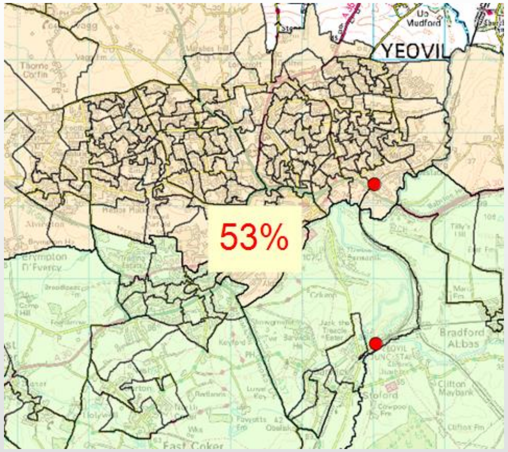
Catchments in reality

- 2km circular catchments account for 57% of trips, 0-20% for some stations (Blainey & Evens, 2011)

- 53% of trip ends located within nearest station zone-based catchments (Blainey & Preston, 2010)

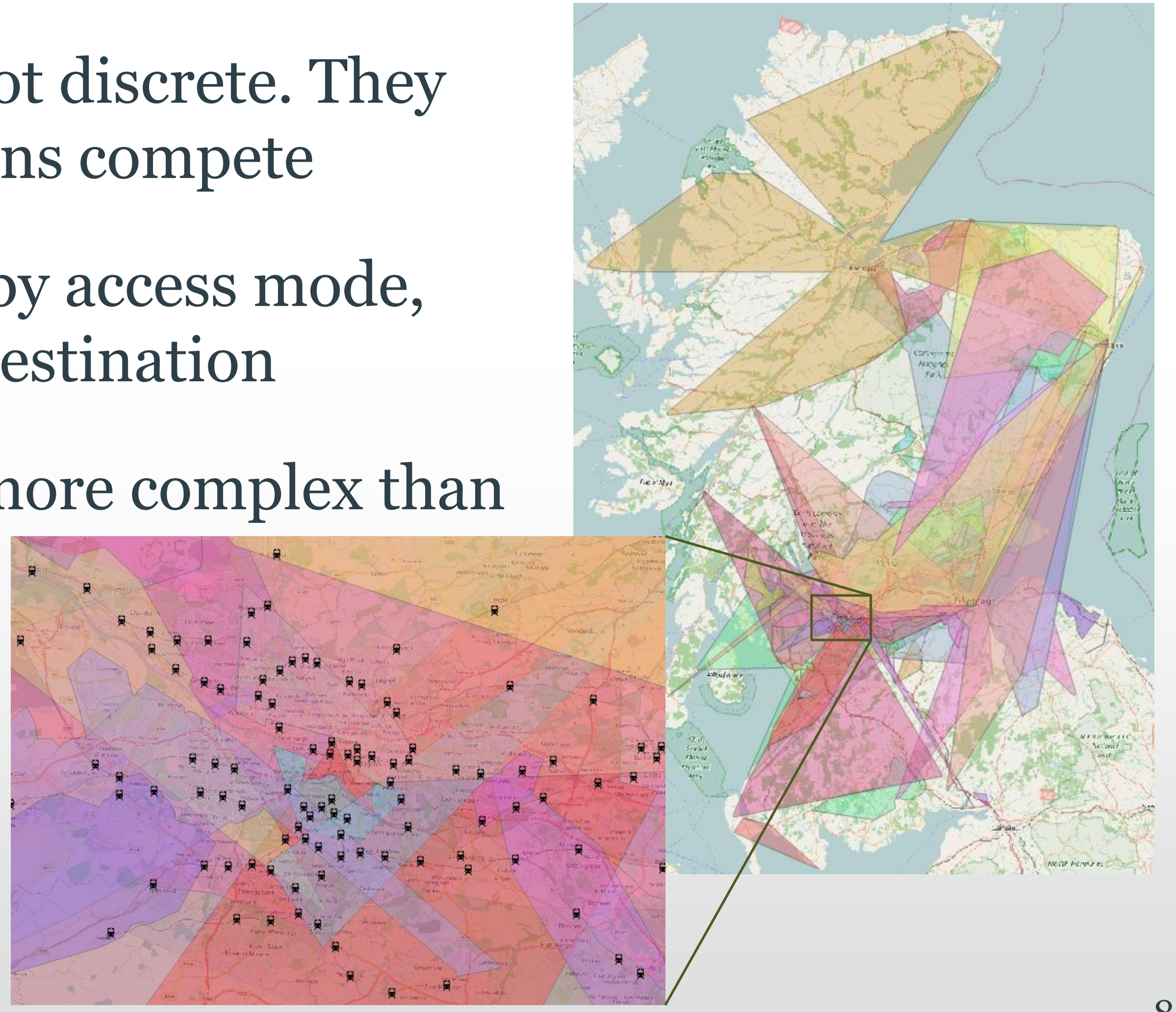
- 47% of passengers in the Netherlands do not use their nearest station (Debrezion et al., 2007)





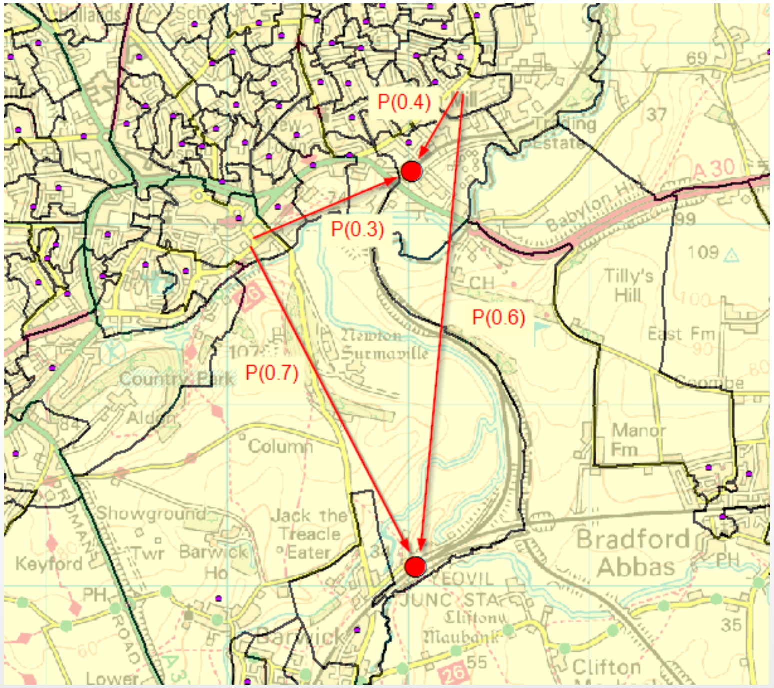
Catchments in reality

- Catchments are not discrete. They overlap and stations compete
- Catchments vary by access mode, station type and destination
- Station choice is more complex than definitions allow



Probability-based catchments

- For each zone calculate the probability of each competing station being chosen
- Allocate zonal population to each station based on the probabilities
- Develop transferable station choice models



UNIVERSITY OF Southampton

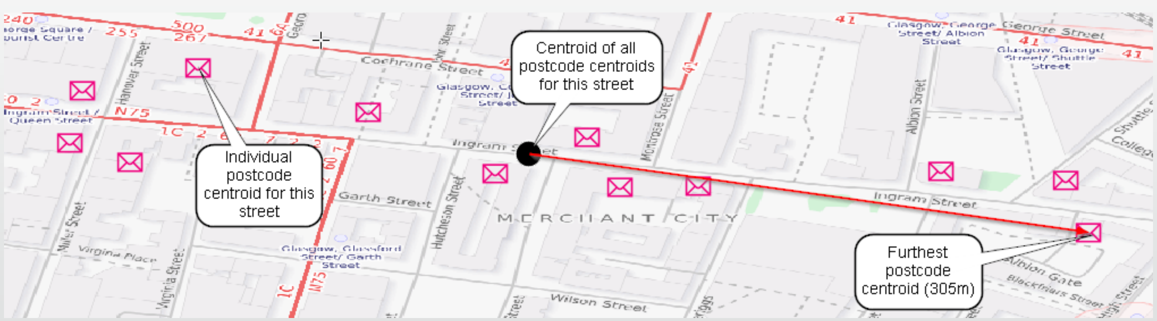
Data considerations

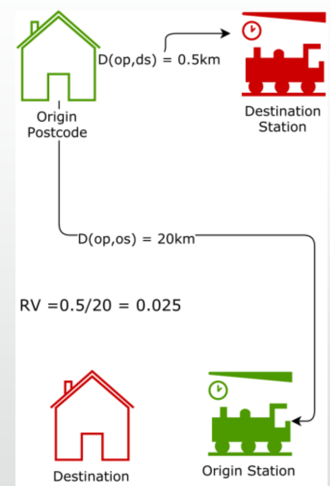
Data – passenger surveys

- Welsh Government (WG) 2015
  - South East Wales (Cardiff, Newport, Valleys) and Swansea
  - 7,000 observations
- LATIS, Transport Scotland 2014 & 2015
  - All of Scotland, concentrated in Central Belt
  - 50,000 observations

Data – cleaning and validation

- Missing unit-level postcodes
  - address matching and location estimation
  - OS AddressBase (28 million addresses)
- Automated trip validation
  - excessive access/egress legs
  - illogical trips (reversed or back-track)





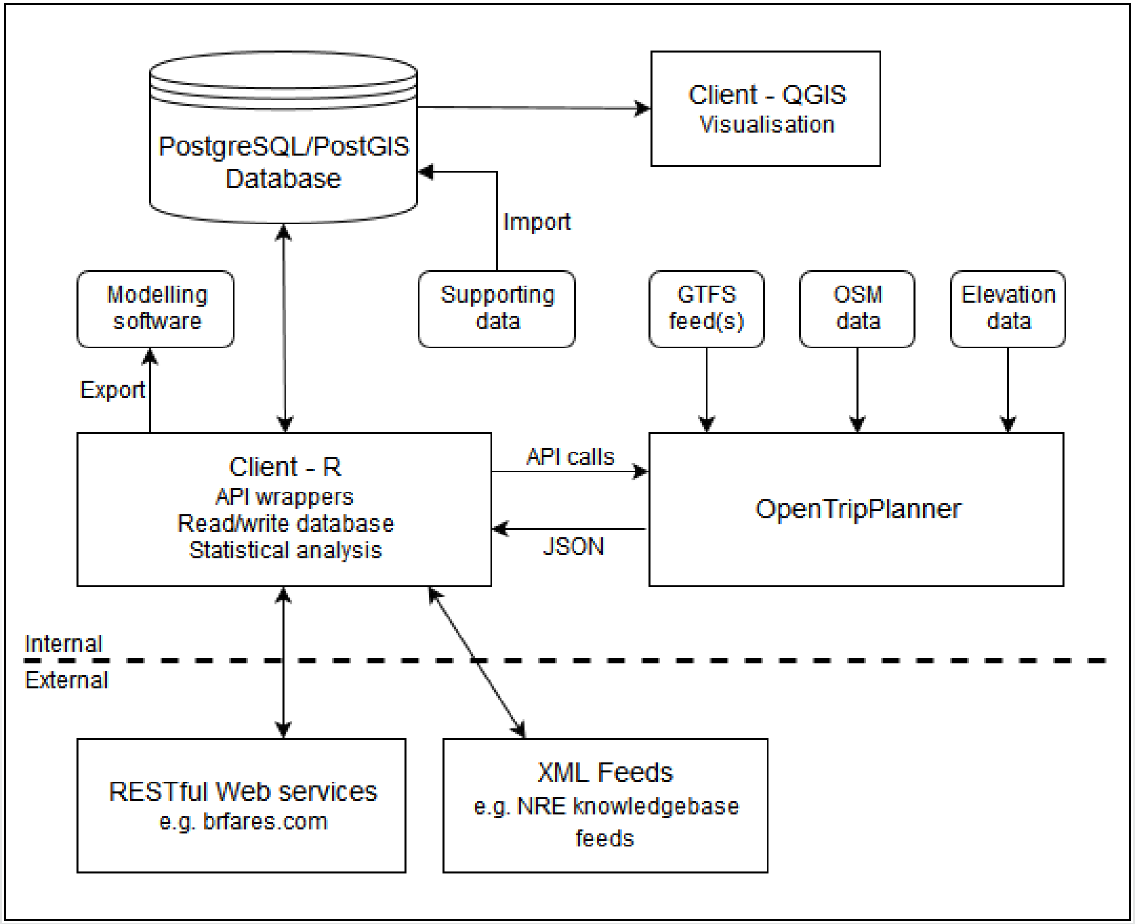
Data – explanatory variables

Access journey	Origin station facilities/services	Train leg
• Drive distance[1] • Drive time[1] • Walk time[1] • Bus time[1] • Nearest station (y/n)[4] • Difference in bearing[4]	• Staffing level[2] • Service frequency[5] • CCTV (y/n)[2] • Car parking spaces[2] • Free car park (y/n)[2] • Toilets/Ticket machine[2]	• Total duration[1] • On-train time[1] • Waiting time[1] • No. of transfers[1] • Fare[3]

Data sources:

1. OpenTripPlanner (bespoke route planner - OpenStreetMap/Transit schedules)
2. NRE Knowledgebase XML Feed
3. BR Fares website
4. Derived from data
5. GTFS feed for GB rail services

Data - Automatic processing framework



Young, M. (2016). *An automated framework to derive model variables from open transport data using R, PostgreSQL and OpenTripPlanner*. Paper presented at 24th GIS Research UK Conference.

Models

Model details

- Discrete choice models – RUM theory
- Choice set varies by individual, defined for each origin
  - 10 nearest stations by drive distance plus nearest major station (97% of observed choice)
- Multinomial logit (MNL) and random parameter (mixed) logit (RPL)

	No. of choice situations	No. of cases	Av. choice set size
LATIS	9,367	97,838	10.44
WG	5,680	59,833	10.53

Model results – MNL

Model	Significant variables	LogLik	Adj R^2
LATIS TE17	• Nearest station • Mode-specific access time • Staffing level • Train frequency • CCTV • Parking space (car mode) • Free car park (car mode) • Ticket machine • Toilets	-6764	.69
WG TE17	As above	-3733	.72
LATIS FM2	TE17 plus: • On-train time • Waiting time • Bearing difference	-5243	.76
WG FM2	As above	-3247	.76

Model results – RPL

- Initial models determined that access time parameters had significant standard deviation (using a log normal distribution)

Model	Specified as random parameters	Significant non-random parameters	LogLik	Adj R^2
LATIS RPL2	• Mode-specific access time	• Nearest station • Staffing level • Train frequency • CCTV • Parking space (car mode) • Ticket machine • Toilets	-6553	.70
WG RPL2	• Access time (walk, bus and car modes)	• Nearest station • Cycle access time (cycle mode) • Staffing level • Train frequency • CCTV • Parking space (car mode) • Free car park (car mode) • Ticket machine	-3649	.73

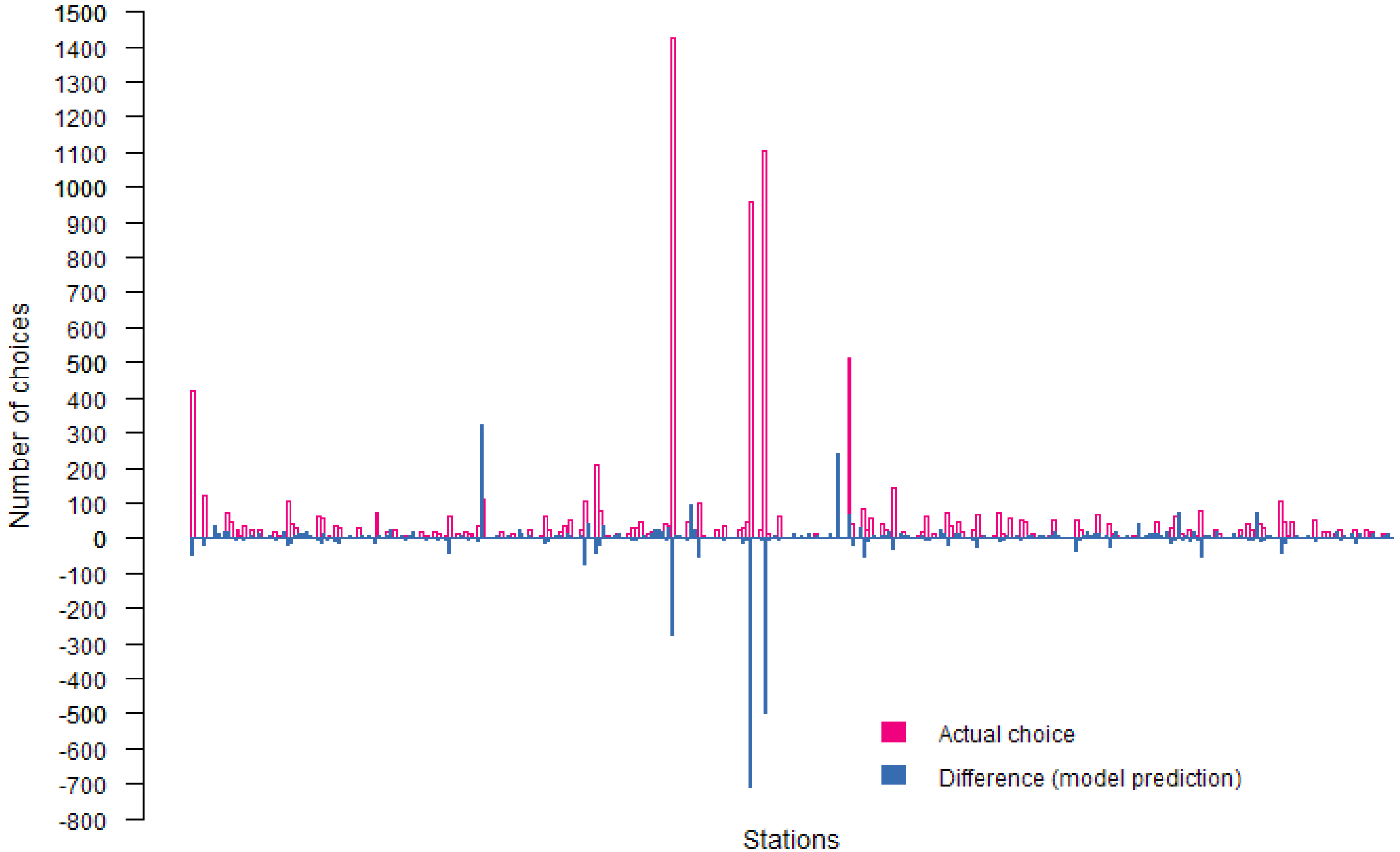
Model appraisal

Predictive performance

- For each station:
  - difference between number of times actually chosen and sum of probabilities from the model
- For entire model
  - absolute difference as % of total choice situations (lower is better)

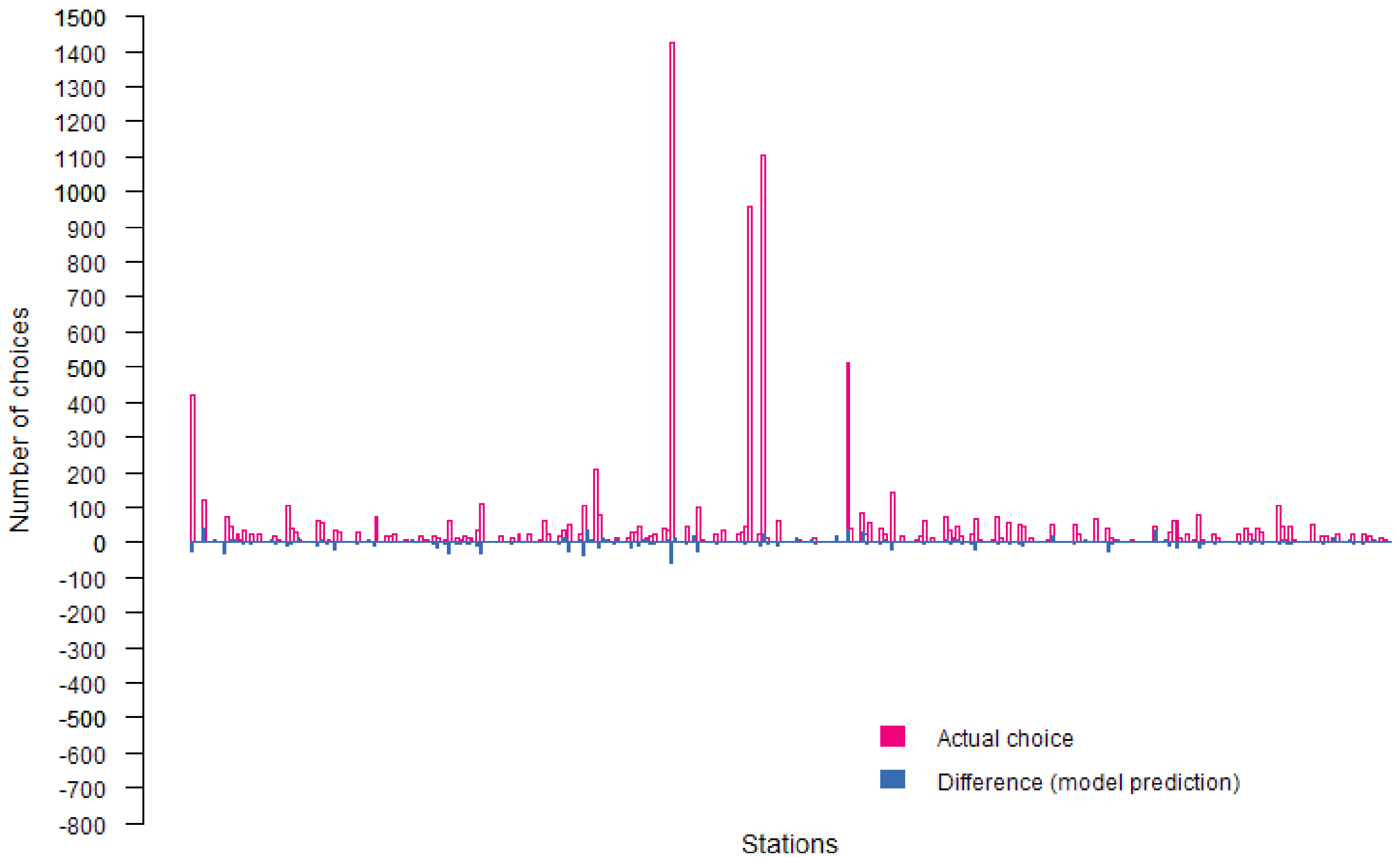
Model	Predictive performance (abs. diff. as % of total choice situations)	
	LATIS	WG
Base model (prob. nearest = 1)	50.91	40.99
TE17	23.53	27.35
FM4	14.61	22.35
RPL2 (TE)	23.58	25.85
RPL4 (FLOW)	n/a	21.13

Predictive performance



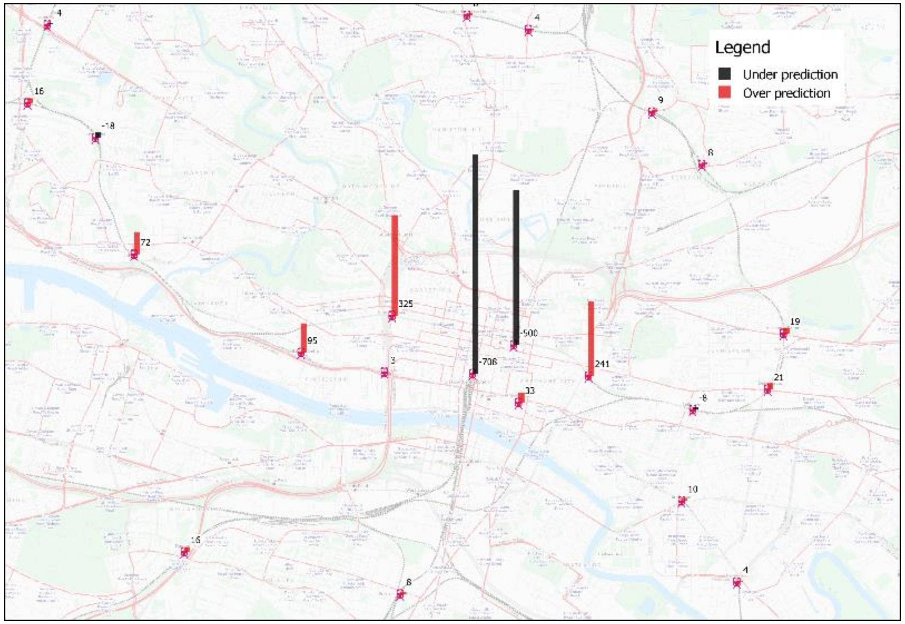
LATIS - base model (Pr. Nearest = 1)

Predictive performance

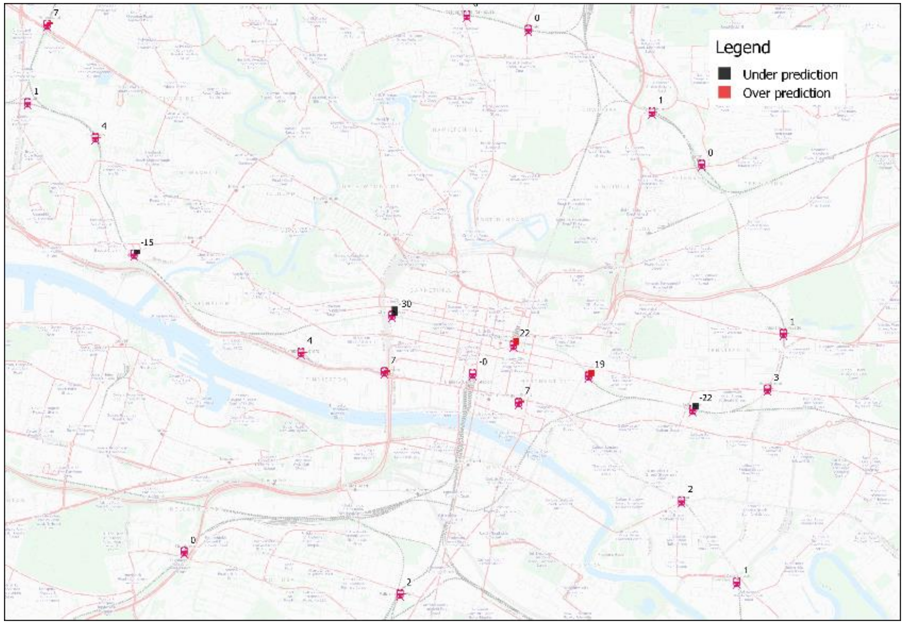


LATIS – FM1 model

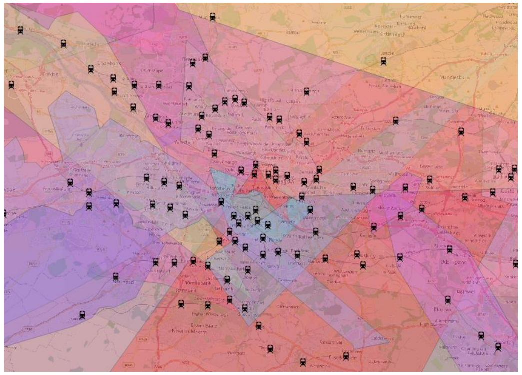
Predictive performance



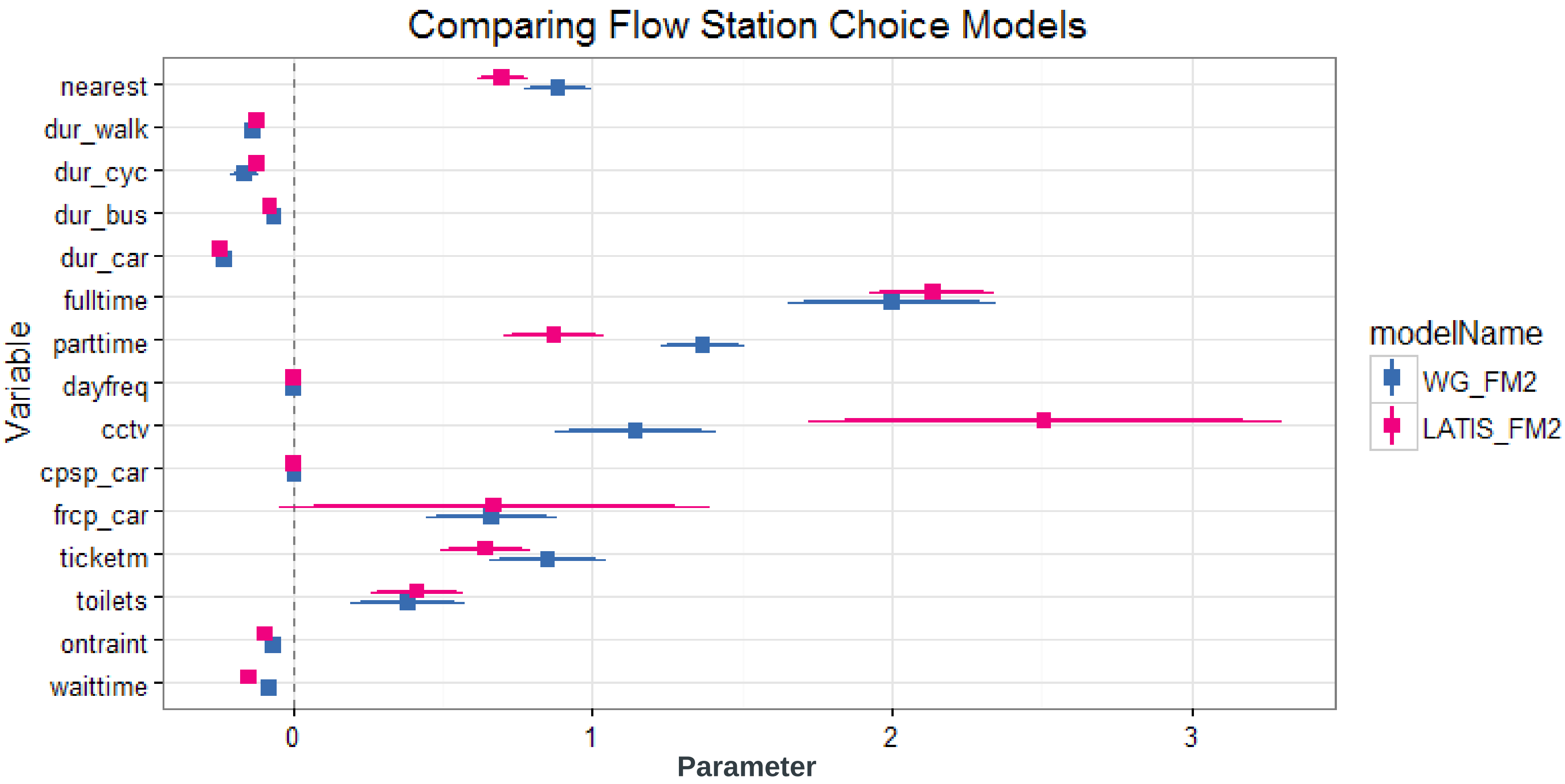
LATIS - base model (Pr. nearest = 1)



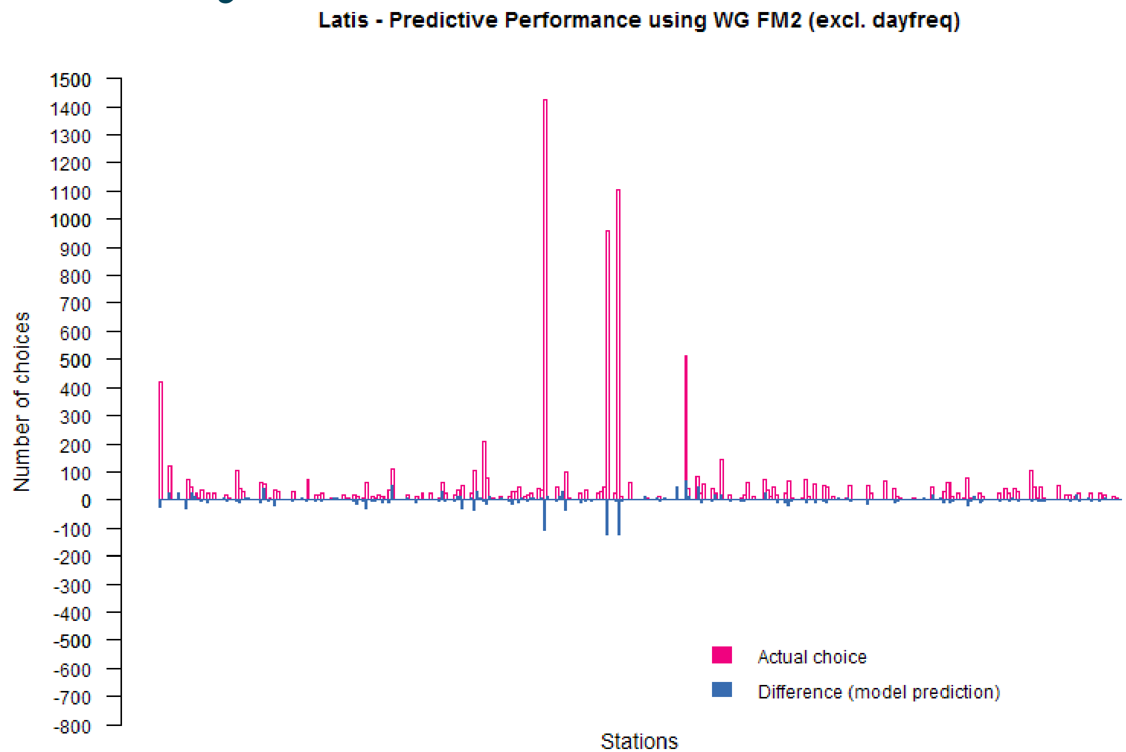
LATIS – FM1 model



Transferability

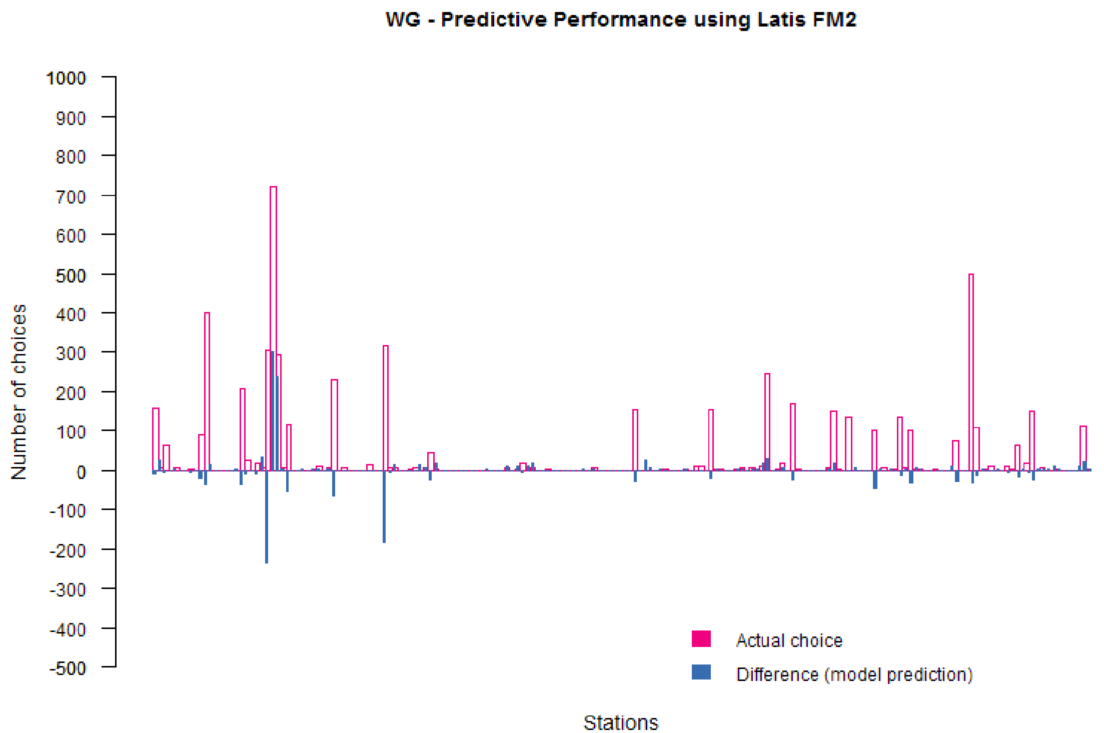


Transferability



Model	Predictive performance	
	LATIS	WG
Base model (prob. nearest = 1)	50.91	40.99
TE17	23.53	27.35
FM4	14.61	22.35
FM2 (model calibrated on other dataset)	$\boxed{20.16}$	34.80

Transferability



Model	Predictive performance	
	LATIS	WG
Base model (prob. nearest = 1)	50.91	40.99
TE17	23.53	27.35
FM4	14.61	22.35
FM2 (model calibrated on other dataset)	20.16	34.80

Conclusions & future work

UNIVERSITY OF Southampton

Conclusions

- Station choice models calibrated using two independent datasets
- Models predict station choice substantially better than a base model that assumes nearest station is always chosen
- Reasonable correspondence in parameter estimates across the two datasets - problematic variables (e.g. CCTV)
- RPL model probably not worth the added complexity, but need to address proportional substitution behaviour of the MNL model

Future work & industry implications

- Main focus to integrate station choice models into trip end and flow rail demand models. When completed this should:
  - improve the models used to assess proposals for new railway stations
  - enable better forecasting of the effects of changing service patterns (e.g. open access services); and
  - provide a methodology that can be incorporated into the industry Passenger Demand Forecasting Handbook

UNIVERSITY OF Southampton

Any Questions?

The author wishes to thank the Welsh Government and Transport Scotland for providing passenger survey data; Paul Kelly for permission to use the brfares.com API; and Ordnance Survey Ltd for providing AddressBase data. This work forms part of a PhD funded by EPSRC DTG Grant EP/M50662X/1.