Visual Analysis of Reactionary Train Delay from an Agent Based Model

Aidan Slingsby[1], Jonathan Hyde[2] and Cagatay Turkay[1]

[1]giCentre, Department of Computer Science, City, University of London, UK
[2]Risk Solutions

Abstract

We design and apply interactive visualisation to help understand reactionary (knock-on) delay in trains using a set of 'Monte-Carlo' style Agent Based Model runs, and to help train operating companies design interventions. Our interactive graphics indicate the locations where primary delay occurs, where reactionary delay occurs, the types of primary delays, how these vary across the runs and the mechanisms of delay. We continue to work with the train industry to refine these methods. Interactive visualisation still has unexplored potential to help make Agent Based Models more explainable.

CCS Concepts
• ***Computing methodologies*** → *Visual analytics;*

1. Introduction

Reactionary delays [Nat19] are delays caused by the knock-on effects of other delays. They affect many types of scheduled service (e.g. air travel [BHBR99]), but we consider reactionary delay on the railways. On UK railways, whilst primary delay has been fairly constant over the last five years, reactionary delay has increased by a third [Rai17]. It is non-linear [CFE*14] and difficult to understand because it depends on other trains moving around a constrained network with limited opportunities to overtake, and to which primary delays (e.g. doors jamming or signal problems) can happen at any time. For these reasons, RSSB (Rail Safety Standards Board; funded by the UK railway industry) is funding research to understand and mitigate the effects of this [Rai17]. They funded us to do a feasibility study with Great Western Railway (a Train Operating Company (TOC) that runs trains between London and the south, southwest and west of UK) in which we combine an Agent Based Model with interactive visualisation to help them design interventions to reduce the effects of reactionary delay. Multiple TOCs run trains on shared railway lines, so reactionary delay is often caused to trains from other TOCs. Delays longer than 3 minutes are recorded, attributed with a reason and causing train (if appropriate), enabling relevant fines and compensation to be paid. The "National Rail Attributed Delays" data is openly available and enable these types of studies to be undertaken.

2. Modelling

One approach to studying the causes and consequences of reactionary delay is to simply consider the historical "National Rail Attributed Delay" data, but with the obvious limitation that we are restricted to situations that have already happened. Statistical models (e.g. [MMSTS15]) that establish statistical relationships between primary delays and reactionary delays would enable reactionary delays to be predicted from primary delays. However, because reactionary delay is driven by the physical mechanism of interacting trains, we opted for an Agent Based Model (ABM) approach [RG11] that has been shown to be feasible [HL09] to model the process explicitly and to help produce explainable outputs.

Our ABM models individual trains moving through the network, based on timetabled times, the railway network, the places where trains can wait (platforms and berths) and the places where trains can be overtaken (additional lines). To these, we add random primary delays (damaged rails, delays due to passengers, delays due to staff), sampled from the "National Rail Attributed Delay" data. We also characterise 'sub-threshold' delays of less than three minutes (not captured in the historical data), using separate movement tracking data from Great Western Railway. For this feasibility study, we used the route between London and Bristol. For each model run, we add random primary delays (based on recent historical data) and the model produces the resulting reactionary delays. Within a set of multiple runs, we consider each run to be an equally valid outcome and the distribution of runs as characterising the variability of delays in the railway. The model generates primary delays and the resulting reactionary delays, with locations, times and train IDs, per model run.

3. Interactive visualisation

Each model runs produces thousands of primary delays (drawn from the recent historical distribution of delays) and thousands of

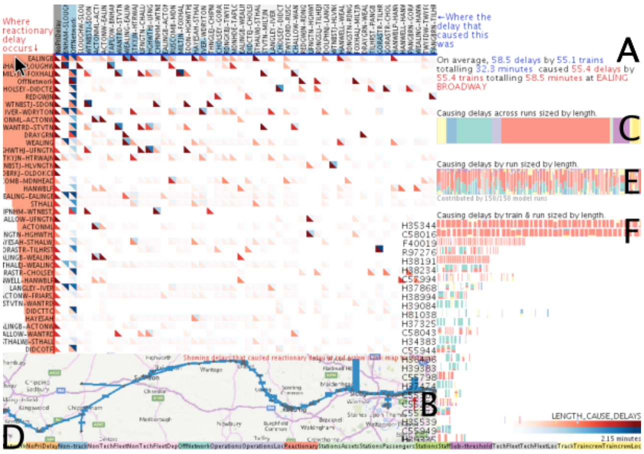

Figure 1: *Matrix showing primary delay as columns (blue) and reactionary delay as rows (red), with stacked barcharts showing the relative delay minutes of causing locations (legend at D).*

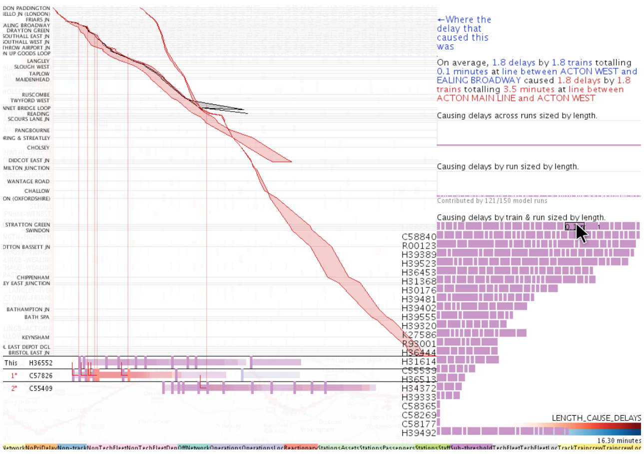

Figure 2: *Delay mechanism for the selected train & run, indicated as a grey shape on time (x-axis) and location (y-axis) graph, with red shapes representing trains subject to reactionary delay.*

reactionary delays (that result from the train movement model). The research challenge is to provide the modelling results in a form that can be used by TOCs to design interventions, but also to enable them consider the sensitivity and robustness of the modelling [FVPS13]. Interactive visualisation has the potential to do this. We depict *variability* between model runs so that TOCs can choose to prioritise locations with more *certain delays* or prioritise locations where *rarer long delays* (lost when averaged) are problematic. Because we model the *mechanism* by which individual trains are delayed, we can provide explanations for the delays. We designed interactive visualisation to help TOCs design interventions to reduce reactionary delay by: (a) identifying pairs of locations where delay causes reactionary delay; (b) showing variability between model runs; and (c) instantly providing graphics that explain delays through the physical mechanism.

Association between *causing* and *reactionary* location. Coloured matrices have long been used to summarise characteristics of object [WF09] or location (e.g. [AA08]) pairs, sorted in some meaningful way [BBW83]. In Fig. 1, *columns* (blue) of the matrix are locations (a station or between two stations) at which primary delays happen. *Rows* (red) are the locations at which associated reactionary delay happens. We split each cell diagonally to show delay characteristics of both locations. The darkness [HB03] of the *blue triangle* indicates the average primary delay minutes (across all model runs) at the source of the primary delay (column). The darkness of the *red triangle* indicates the resulting reactionary delay minutes at the resulting location (row). Where a cell's blue colour is lighter than its red colour, a comparatively *short* primary delay leads to *longer* reactionary delay, locations where reactionary delay might be *reduced* more easily. The text on the right (A) summarises the delay situation of the reactionary location indicated by the mouse pointer, to help interpretation. Both rows and columns are sorted by *reactionary* minutes that result, so the topmost/leftmost are the locations where interventions should be focussed. Row/column locations are indicated on the map (B) with a red/blue arrow, helping put the location in its geographical and rail network context.

Variation between model runs. The stacked barchart (C) shows the average ratio of primary delay types *across the model runs* (legend at D), with the 'reactionary' (red) causing most of the reactionary delays, followed by 'off network" (green). *Variation between model runs* is depicted in the 150 adjacent stacked barcharts (E), one for each model run. Bar width indicates delay minutes and bar height indicates ratio of delay minutes by primary delay type.

Explaining delays. When designing interventions, it is important to know which trains are involved and why. Since we model the delay mechanism, we can show this to *explain* delays at locations on-demand. For the location or pair of locations indicated with the mouse, the top *causing trains* are listed (F), where each segment represents a model run, coloured by primary delay type. Fig. 2 illustrates what happens when the mouse is moved over a specific train and run. In this case, all are affected by 'sub-threshold' delay (purple). The matrix is replaced with a time (x-axis) and location (y-axis, London to Bristol) graphic. This is a modified version of Marey's graphical train schedule [Tuf01], in which the left edge of the shaded area is the timetabled time, the right edge is the actual time (width is amount of delay). The grey shaded area is the train indicated by the mouse, labelled as "this" at the bottom. The red shaded areas are trains delayed by this train, with vertical red lines indicating when the reactionary delay occurred, and the "1°" and "2°" labels indicating first- and second-degree reactionary delay.

4. Conclusions

We show the feasibility of using interactive visualisation with ABMs to help understand reactionary delay and help design interventions. Our interactive graphics indicate important locations, variability between model runs, and explanation through specific mechanisms of delay. We continue to work with TOCs to refine these methods. Interactive visualisation still has unexplored potential to help address some of the open research challenges of ABMs, such as sensitivity analysis, verification and validation [FVPS13] and make ABMs more explainable.

References

[AA08] ANDRIENKO G., ANDRIENKO N.: Spatio-temporal aggregation for visual analysis of movements. In *2008 IEEE Symposium on Visual Analytics Science and Technology* (Oct 2008), pp. 51–58. doi:10.1109/VAST.2008.4677356. 2

[BBW83] BERTIN J., BERG W. J., WAINER H.: *Semiology of graphics: diagrams, networks, maps*, vol. 1. University of Wisconsin press Madison, 1983. 2

[BHBR99] BEATTY R., HSU R., BERRY L., ROME J.: Preliminary evaluation of flight delay propagation through an airline schedule. *Air Traffic Control Quarterly 7*, 4 (1999), 259–270. 1

[CFE*14] CAMPANELLI B., FLEURQUIN P., EGUÍLUZ V., RAMASCO J., ARRANZ A., ETXEBARRIA I., CIRUELOS C.: Modeling reactionary delays in the european air transport network. *Proceedings of the Fourth SESAR Innovation Days, Schaefer D (Ed.), Madrid* (2014). 1

[FVPS13] FILATOVA T., VERBURG P. H., PARKER D. C., STANNARD C. A.: Spatial agent-based models for socio-ecological systems: Challenges and prospects. *Environmental modelling & software 45* (2013), 1–7. 2

[HB03] HARROWER M., BREWER C. A.: Colorbrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal 40*, 1 (2003), 27–37. 2

[HL09] HWANG C.-C., LIU J.-R.: A simulation model for estimating knock-on delay of taiwan regional railway. *Proceedings of the Eastern Asia Society for Transportation Studies 2009* (2009), 213–213. doi:10.11175/eastpro.2009.0.213.0. 1

[MMSTS15] MARKOVIC N., MILINKOVIC S., S. TIKHONOV K., SCHONFELD P.: Analyzing passenger train arrival delays with support vector regression. *Transportation Research Part C: Emerging Technologies 56* (07 2015). doi:10.1016/j.trc.2015.04.004. 1

[Nat19] NATIONAL RAIL: *How a delay to services in one area can affect trains elsewhere in the country*. Tech. rep., 2019. URL: https://www.networkrail.co.uk/running-the-railway/looking-after-the-railway/delays-explained/knock-on-delays/. 1

[Rai17] RAIL SAFETY AND STANDARDS BOARD (RSSB): *Call For Research: Data Sandbox: Improving Network Performance*. Tech. rep., 2017. URL: https://dkf1ato8y5dsg.cloudfront.net/uploads/15/81/data-sandbox-improving-network-performance.pdf. 1

[RG11] RAILSBACK S. F., GRIMM V.: *Agent-based and individual-based modeling: a practical introduction*. Princeton university press, 2011. 1

[Tuf01] TUFTE E. R.: *The visual display of quantitative information*, vol. 2. Graphics press Cheshire, CT, 2001. 2

[WF09] WILKINSON L., FRIENDLY M.: The history of the cluster heat map. *The American Statistician 63*, 2 (2009), 179–184. URL: https://doi.org/10.1198/tas.2009.0033, arXiv:https://doi.org/10.1198/tas.2009.0033, doi:10.1198/tas.2009.0033. 2