Predicting and mitigating small fluctuations in station dwell times

James Pritchard[a], Jason Sadler[b], Simon Blainey[a], Ian Waldock[b], Jeremy Austin[b]

[a] *Transportation Research Group, University of Southampton, Boldrewood Campus, Building 176, Southampton SO16 7QF, United Kingdom*
[b] *GeoData Institute, University of Southampton Highfield Campus, University of, Southampton SO17 1BJ, United Kingdom*

Abstract

On busy railway networks, exceeding station dwell times by only a few seconds can adversely affect overall network performance. However, while these performance impacts are well known, the causes of small dwell time perturbations are not widely understood (or are not widely communicated at an operational level) and exhibit a high level of spatial and temporal variation. A lack of information and understanding makes it difficult to implement effective mitigation measures to reduce the occurrence and impact of such delays.

For this paper On Train Monitoring Recorder (OTMR) data were obtained for a large number of services over a 14-month period, which included the wheel stop and wheel start timings associated with station stops at a one second resolution. These were combined with other relevant data sources in order to investigate small fluctuations in station dwell time. An interface for communicating these variations to railway operating staff was developed, along with models to predict future dwell time fluctuations, potentially enabling mitigation measures to be implemented.

Keywords: station dwell; delay; modelling

1. Introduction

In any mode of public transportation, station dwell time is a key parameter of system performance and service reliability [14]. Variations in station dwell times can have a significant effect on the capacity of a rail network; even small delays of a few seconds can adversely affect overall network performance. As with most other delays, as the network becomes busier they will have an increasing impact on the punctuality of other services, leading to growing levels of overall reactionary delay. Variations in dwell times can arise from a range of factors, ranging from problems with door mechanisms to issues associated with fast-changing weather conditions; although unplanned, such variation is not always random and even some that is can be understood and hence managed [5]. Identification of 'hotspots' (stations and/or services which may be prone to dwell time variation which leads to delays) and an understanding of the underlying cause(s) are important first steps in developing effective mitigation strategies. Possible mitigation strategies include long-term planning (e.g. revising the timetable to make it more robust), short-term operational interventions (e.g. ensuring that staff are prepared to manage an upcoming station dwell as efficiently as possible) and improvements to station infrastructure (e.g. providing better weather protection to help encourage passengers to spread themselves along the length of the train).

On the UK railway network, current delay attribution data only provide causal information on delays of three minutes or more, even though shorter initial delays can be quickly compounded across the network. These delays and their knock-on impacts can have a significant impact on the quality of the rail service which is offered to passengers, and therefore on passengers' perceptions of the quality of service offered by rail relative to other modes of transport. A lack of information and understanding can make it difficult to implement effective mitigation measures to reduce the occurrence and impact of such delays. Given the strong case for encouraging mode shift to rail as part of efforts to reduce the environmental impact of transport, there are clear societal benefits from the development of methods to understand and reduce fluctuations in station dwell times.

The aim of the research described in this paper was to develop models which are capable of making real time predictions of expected variations in station dwell time, along with an interface for communicating these variations to railway operating staff so that they can implement appropriate mitigation measures. In order to achieve this aim, the research addressed the following three objectives:

1. Develop a model explaining the causes of small fluctuations in station dwell time based on a range of datasets.
2. Explore the use of this model to forecast future fluctuations in station dwell time, in real time.
3. Develop a system for alerting railway operating and planning staff to the potential for extended dwell times, and suggesting potential mitigating actions.

In order to achieve these objectives, data were obtained from a UK rail operator, covering the majority of their fleet over a 14 month period. Data included records of wheel-stop and wheel-start timings (down to the second) for station stops. This paper begins with some background, including a review of definitions of station dwell time, before describing the methodology used in this case to analyse the data and develop appropriate models. Results from the development of regression models (used predominantly to identify the important factors which influence station dwell time) and from classification models (designed to support tools which could be provided to operational staff) are presented.

2. Background

2.1. Defining station dwell time

There appears to be a consensus that station dwell time can be defined generally as the time for which a train remains stopped in a railway station for the prime purpose of allowing passengers to board and/or alight (e.g. [7], [17]). It is widely acknowledged that scheduled dwell time can be divided into several components; they typically include the time required to open the doors, time for passengers to alight and board, and the time required to close the doors and dispatch the train.

A lot of research has focussed on the passenger alighting and boarding phase, with the phases either side being seen to comprise relatively fixed components (e.g. [2]) or simply thought of as 'lost time' [7]. However, variations in these phases can also contribute to sub-threshold delays. Causes of such variation can include mechanical issues (e.g. poorly performing door mechanisms) and the actions of personnel involved in the processes. Depending on the context and the mode of operation, the actual departure time could depend not just on the reactions of a driver [6], but on the actions and reactions of a guard, platform staff and signallers.

Scheduling can also have an impact on station dwell time. Strategies for improving punctuality can include the use of buffers such as increased (scheduled) station dwell time [11]. In some literature, buffer time is explicitly included as a distinct component of station dwells (e.g. [6], [2]). Planned buffer times are invariant [2] (for a given schedule), but the actual impact on dwell time overall is less certain. Although a corresponding reduction in delays might be expected from an increase in scheduled time, it is not necessarily the case in practice: It is widely observed that if more time is allowed for an activity, then the activity itself often takes longer to complete [11][1], and there is some evidence that increasing the time for boarding and alighting may induce undesirable passenger behaviour. For example, it may give the perception that deliberately holding doors open is relatively inconsequential, leading to the recommendation that "operators wishing to reduce the occurrence of user-induced delay for scheduling or user safety purposes could consider minimising the time allotted for boarding and alighting" [18].

Train and station design can also influence dwell time, particularly the boarding and alighting phase. Interior features of the train such as aisle width, the presence of 'perch seats' in the vestibule and the provision of luggage racks have all been shown to have some impact on boarding and alighting flows [10], as have door width, spacing between doors along the train and the step between the platform and the train [4]. Platform width is also a consideration, whilst the location of station entrances, canopies and customer information boards can lead to uneven distribution of passengers along a platform [10]. The duration of a station stop (or at least, the boarding and alighting phase) can be determined by the busiest door along a train [4] and uneven distribution of passenger movements and passenger loads between carriages is known to impact dwell time estimation models [21]. Train design may also have some impact on other phases of station dwell: For example, the time taken to open and close the doors would be expected to vary between different rolling stock designs. Much of the literature has focussed on how train design impacts the movement of passengers, but there are aspects (e.g. the placement of door release controls) which could potentially have an impact on tasks performed by operating staff.

Some of the key components of dwell time can be subdivided - for example, the time required to open the doors includes both mechanical processes and other human factors, such as the behaviour of the both the door operator and the passengers. In other cases, some of the components overlap; reaction time could also be a factor in the time required to open or close the doors, whilst any buffer time could be absorbed across different phases. An advantage of defining the phases quite broadly is that they can equally be applied to automated systems where some of the human factors are not relevant.

This paper considers station dwell time as everything between wheel stop and wheel start. Although the data provided were not sufficient to analyse any of the individual components within this time, having an awareness of the different elements is nonetheless important when seeking to understand the fluctuations in dwell time and the reasons for them.

2.2. Factors which influence station dwell time

There are a number of factors which influence station dwell time, and they may be grouped as follows:

- Passenger numbers
- Impact of train design

- Train operation (this includes the actions of the driver and other staff, such as the guard, and may be influenced by external operations, including signalling).
- Network operation (including signalling rules and clearances)
- Station operation (including dispatch staff)
- Passenger characteristics (including their age, the amount of luggage and their familiarity with the journey) and passenger behaviour
- Impact of station architecture on passenger movements
- External factors (including weather, traffic conditions and system failures)
- Train scheduling (including the use of buffer time, and any co-ordination between services or rules about connections)

Although these groupings are largely based on existing literature [9][15], they should not be taken as definitive. For example, it could be preferable to separate the "train operation" category in to human factors and mechanical factors. Some factors may span more than one grouping (e.g. "door opening and closing time" is partly a characteristic of the train itself and partly influenced by the operating staff and the passengers), whilst others may be seen as indirect (e.g. poor weather may impact the behaviour of the passengers on the platform, especially at stations where shelter is limited).

The groupings suggested here also implicitly consider a single station-stop in isolation, although aspects of the train service as a whole could be considered as "external factors." One paper considered here was a notable exception, focussing instead on predicting departure delays from arrival delays [3], including late arrivals of connecting services. Given that the general definition of station dwell time includes reference to the time for which a train is stopped, it does not explicitly take in to account a late arrival, but it may nonetheless be an influencing factor. In the case where multiple train services are co-ordinated, transfer connections can be of particular importance [3]. Holding trains for connecting services leads to an increase in station dwell time, although this is not currently standard practice on the British railway network (with some rare exceptions, for example, to avoid leaving passengers stranded en route if it is the last train at night).

2.3. The value of research in to station dwell time

It has been suggested that station dwell time is one of the biggest constraints on maximising rail capacity [15], and there is often a trade-off between efficiency of operations and punctuality [7]; a tight dwell time can be a source of delay, whilst longer dwell times can lead to extended journey times and a low utilisation of platform track capacity [21]. Station dwells are an operational constraint because a train is not typically permitted to enter a platform until the preceding train has departed; when a rail system is operating close to its capacity, small irregularities in service can lead to delays [8], and these can propagate throughout the network. The risk of delays can be minimised by adding buffer time (either through specific recovery time allowances, or 'generous' timings throughout the journey), but this impacts capacity. Excessive dwell time is inefficient. It could be argued that variation in actual train departure times is inevitable, even if the scheduled dwell time includes some buffer time [3], but there is consensus in the literature that understanding the causes of variation and working to optimise dwell times is beneficial. From an operational perspective, "dwell time studies give insight into the travel time and headway variations and can produce effective timetables" [17]. In addition to being useful for long-term timetable planning, the prediction of train dwell times at stations is important for real-time rescheduling (for example during disruption) [9], whilst "analysing the interaction between dwell time and the delays of crowded rail transit lines is extremely useful toward effectively managing passengers during delays" [6]. It is clear that it is not just rail operators who stand to benefit from a better understanding; passengers also benefit from increased capacity and reduced delays. Indeed, the ability to achieve tight dwell times on a consistent basis is "critically important for achieving the planned capacity, delivering the right passenger experience and end-to-end journey times" [19].

Much of the research into station dwell times to date has focussed on passenger flow; given that station dwells exist to facilitate passenger movements (in the form of boarding and alighting), this focus is perhaps not surprising. However, a key limitation of some of the models, including the widely referenced Weston Model [20],is their dependence on data about passenger numbers, often in some detail (including figures for boarding and alighting at every door). Passenger counting data are becoming increasingly widespread, although not every system used to capture passenger numbers is capable of providing the required levels of accuracy and granularity for dwell time modelling. Some systems (for example, those which infer passenger numbers from mobile network data) only make estimates of whole train occupancy, whilst others (such as those which use on-train electronic weighing equipment) can estimate carriage occupancy but do not provide passenger flow at individual doors. One of the biggest challenges, however, is ensuring that crowding information are provided in a timely enough manner for key decisions to be taken [13]. This was a particularly important consideration here, given the aim of developing a system which can alert staff to the potential for extended dwell times. Some trains in the UK are now equipped to provide passenger counts in real-time, and the proportion of such trains may reasonably be expected to increase. However, the available data are not currently sufficient for widespread real-time decision support tools (especially given that some operators are only fitting a subset of trains in their fleets with the required systems).

A key study in the Netherlands [9] has shown that it might be possible to predict station dwell times without detailed information about passenger flow, using a range of operational metrics (such as dwell time of previous services) which may be used as a proxy for e.g. high passenger numbers. Compared with other work, this study also has the added benefit of being an example of a non-metro system with less homogenous stations and rolling stock. There is a need to understand whether such models could be more widely applicable, and to ascertain their usefulness for e.g. improving operational practices. This paper aims to address both of these issues.

For this paper, data were obtained for a major UK rail operator, including On Train Monitoring Recorder (OTMR) data which recorded wheel-stop and wheel-start timings to the second for each scheduled station dwell. The data were analysed and regression models were developed, based on the possible predictors used in the study by Li et al. [9], in order to understand the fluctuations in station dwell time. Classification models were also developed using a Random Forest methodology. A visualisation application was also developed (Section 4.4), in order to help operational staff visualise the issues. The ability of the different models to provide useful future predictions (so that appropriate mitigation measures could be taken) was tested.

3. Methodology

Data were obtained from several sources and were stored in a PostGreSQL server with PostGIS geospatial extensions. A range of software tools were used to access, process and visualise the data and to develop statistical and predictive models: these included database tools such as DataGrip, programming languages such as Python and R, desktop Geographic Information System (GIS) tools such as QGIS and ArcMap and interactive web tools for displaying dynamic charts and maps.

3.1. Data sources

Data for this project included:

- The Network Rail (NR) Rail Infrastructure Network Model, used to monitor individual assets. (Data provided included additional lookups from Network Rail, and involved some internal processing)
- Rail reference data from the Open Rail Data Wiki [12]
- Signalling and train movement data covering the first four months of 2018.

- Selected OTMR data from a UK Train Operating Company (TOC), including wheel-stop and wheel start timings at each station stop. The data covered the majority of the operator's fleet from April 2017 until August 2018.
- Additional data from the TOC, including train allocation data, in order to ensure that the OTMR data could be linked to a particular train on a particular service on the network

It was initially planned to incorporate historic and near-real time weather and environmental data; however, the cost of receiving data at a sufficient spatial resolution was prohibitive.

3.2. Initial regression models

Regression models for the variation of actual dwell time from scheduled dwell time were developed in Python. The data obtained for this project were concerned with train operation, not passenger flow, and the initial feature set (Table 1) was chosen accordingly, based on existing research [9]. A backwards linear regression method was used, which automatically eliminated any variables not found to show significance (using a threshold of $p = 0.01$).

Table 1: Initial features selected for the models

Feature	Related influence(s) on dwell time	Expected relationship to dwell time
The dwell time (in seconds) of the same service the previous week	Regular passenger numbers or other regular factors	Positive

Boolean variable denoting whether the service in question was at a weekend	Passenger numbers (demand varies between weekdays and the weekend)	Negative (reduced weekend demand leading to a reduction in dwell time)
Boolean variable denoting whether the service in question was a peak time train	Passenger numbers (demand varies between peak and off-peak)	Positive (increased peak demand leading to an increase in dwell time)
Boolean variable denoting whether the order of preceding services was as scheduled	External factors leading to operational delays	Positive if disruption has lead to high levels of crowding on the platform. Negative if the schedule contains slack for making up time
The dwell time (in seconds) of the service at the previous station	Passenger numbers (boarding time at the previous station dependent on passenger flow) The train operating the service (any issues with train operation - e.g. slow door closing - will likely be evident at previous stops)	Positive
The dwell time (in seconds) of the service at the second previous station	Passenger numbers The train operating the service	Positive

The difference between the actual time of departure and the scheduled time of departure (in seconds) at the previous station	Passenger numbers (including the possibility that a delay has lead to an increased build up on the platform) External factors leading to operational delays	Positive
The dwell time (in seconds) of the previous service in the same service group	Passenger numbers	Positive
The dwell time (in seconds) of the previous relevant service	Passenger numbers Characteristics of the station	Positive
The train length (number of carriages, based on actual formation data)	Passenger numbers (at each door) Passenger behaviour (e.g. where the passengers are standing along the platform)	Unclear. A longer train might be expected to be busier overall and may take longer to dispatch, but a shorter train may lead to higher passenger numbers at each door.
The length of the preceding train	Passenger behaviour	Unclear

The definition of a “peak” train can vary across between locations, but in this case a peak-time train was defined as any train departing the station in question on a weekday before 10am or between 5pm and 7pm. Service groups were defined by ’headcode’ - a four digit alphanumeric identifier linked to a given service schedule. For the network studied, the first two digits of the headcode are typically assigned to a particular service pattern (for example, ’1Sxx’ headcodes may refer to express services between Station X and Station Y) and the data were grouped accordingly. Busy stations may be served by a number of different service groups heading in the same direction, and the previous relevant service was defined as the previous service in the same direction. Train length data were not contained within the OTMR data and had to be inferred separately from train allocation data supplied by the TOC.

There are a some caveats associated with this initial approach, including:

- In some cases, the previous relevant service and the previous service in the service group are the same thing
- Stopping patterns within a particular service group may not always be the same (especially at the beginning or end of the day)
- The supplied train allocation data was not guaranteed to be accurate. Furthermore, there are some cases of headcodes being assigned to more than one service in a day (where there is enough separation in time and space to guarantee no physical conflict on the network). In most cases, there were sufficient additional data to avoid issues, but there remains some potential for ambiguity when it comes to train formation.

Data were filtered to exclude cases where the (absolute) difference between the actual and the scheduled dwell time exceeded 180 seconds. This was to:

1. Help exclude cases which may be based on erroneous data. For example, data for some journey points suggests an early departure of several minutes or more, which seems excessive even after allowing for a bit of discrepancy between different timetables and data sources.
2. Help exclude cases where unusual external factors may have had a big impact (e.g. widespread disruption, signal failure, train failure or medical emergency).

The threshold of 180 seconds was chosen to reflect the fact that the focus of this project is on delays below the current delay attribution threshold of three minutes; above this threshold, the cause of the delay is already investigated, with associated costs attributed to the responsible parties.

The network operated by the TOC is diverse, and there is currently a lack of data which would allow stations to be suitably characterised and categorized. The factors which may influence dwell time, such as the position of a shelter on the platform, are independent of the usual methods for grouping stations (e.g. by size or by footfall). Hence models were developed for each station and service group separately, in order to provide improved accuracy and minimise the influence of unknown externalities, such as the design of a particular station, on an overall model. Although this limits the generic applicability of any outputs at this stage, such models are always going to be context-dependent. This approach does enable stations and services of particular interest to be identified and perhaps prioritised for further investigation.

3.3. Scenarios chosen for initial regression modelling

Three stations were selected for the initial analysis; they were chosen in order to capture the diversity of services within the region studied and to help generate outputs of relevance to large parts of the wider UK network. Two of the stations lie within Greater London (where the network is congested and sub-threshold delays can be a particular issue), whilst the third is a mainline interchange outside London served by longer-distance 'inter-urban' services. Different service groups were considered in each case, leading to six scenarios in total (Table 2). Service groups are typically bi-directional and the characteristics of each direction of operation may differ. Each scenario therefore used data from a single direction of operation: Table 2 uses the British convention of labelling services towards London as 'Up' and services originating at a London terminus as 'Down.' Scenarios C1 and C2 look at the same service group at the same station, comparing trains in the 'Up' and 'Down' directions respectively.

Whilst dealing with a small subset of stations and services, it was possible to explicitly specify scheduled dwell times, both for the stop in question and for those which preceded it, when selecting the data for analysis. This reduced the impact of non-homogenous stopping patterns and ensured that the models were not dominated by scheduling variations. It also enabled some scheduling variations to be considered explicitly: For Station B, a single service group was analysed, but the peak-time services with additional scheduled dwell (60s instead of 30s) were considered separately.

In addition to excluding those services recorded as exceeding the scheduled dwell time by more than 180s, the data were further filtered on arrival times. Services recorded as arriving at the chosen station three minutes or more ahead of schedule were excluded on the basis that these instances were either likely to be erroneous or as a result of exceptional circumstances (for example, an occurrence of planned engineering work resulting in an amended timetable with very large amounts of buffer). Services recorded as arriving 20 minutes or more behind schedule were excluded, partly to further reduce the potential of erroneous data and partly to exclude instances of extreme disruption. The 20 minute threshold allows for some impacts of perturbed operations to be considered whilst filtering out any instances of major disruption in which both trains and staff could be badly displaced and passenger flows could be exceptionally abnormal.

Table 2 shows some significant differences between scheduled dwell times and observed dwell times. A major cause of this is likely to be scheduling practices. Several routes converge at Station A, and the scheduled dwell times typically include significant buffer time to help services regain time lost en route before continuing along the increasingly congested line towards London. This is reflected in the fact that the dwell times are significantly shorted than scheduled for scenarios A1 and A2, and a significantly high proportion of trains arrived behind schedule (87% of trains in A1 and 72% of trains in A2). The mean dwell time for peak services in scenario A1 is also less than the mean dwell time for off-peak services, which may indicate that more use is made of the buffer time (due to earlier delays). It may also indicate that the operations during the peak time are a bit slicker, perhaps because of a sense of pressure when the system is busier (both in terms of throughput of trains and of people).

Table 2: Scenarios selected for initial analysis

Scenario		A1	A2	B1	B2	C1	C2
Station		A (Non-London)		B (London 1)		C (London 2)	
Service Type		Inter-urban		Suburban		Suburban	
Traction Type		Electric	Diesel	Electric	Electric	Electric	Electric
Direction		Up	Up	Up	Up	Up	Down
Scheduled Dwell (s)	This stop	120	120	60	30	30	30
	Previous stop	90	60	60	30	30	30
	Prev. stop + 1	90	60	60	30	30	60
# of observations	All	992	456	496	1646	2224	2153
	Peak services	371	456	496	254	730	640
	Weekend services	137			608	525	451
Mean actual dwell (s)	All services	87.5	97.6	44.6	39.1	46.5	44.7
	Peak services	85.0	97.6	44.6	38.2	50.0	45.2
	Off-peak weekday services	89.0			37.9	45.8	44.8
	Weekend services	89.1			41.0	42.9	43.8
Dwell std. deviation (s)	All services	34.2	26.7	14.4	15.4	21.8	15.6
	Peak services	30.7	26.7	14.4	11.5	23.7	14.8
	Off-peak weekday services	36.7			15.6	21.6	16.6
	Weekend services	33.8			16.3	18.5	14.2

Stations B and C are examples of stations where train planners work to 45s dwell times whilst the working timetable is cast in 30s intervals. This is achieved in practice by using a mix of 30s and 60s scheduled station stops along the route. The observed dwell times for scenarios B1,2 and C1,2 are broadly consistent with this, but the strategy is problematic when it comes to sub-threshold delays. If the working assumption is that some level of excess dwell (with respect to the timetable) will be absorbed elsewhere on the route, the case for addressing sub-threshold delays is made harder. Scenario C1 is a good example of where there may be insufficient slack in the system (even when assuming a planned 45s dwell and not the scheduled 30s), with 62% of services classed as having lost time at the station (defined as services which did not arrive ahead of schedule, and where the departure delay was greater than the arrival delay). In line with the observed increased mean dwell, this proportion increased to 67% during peak times. The time lost may still be absorbed elsewhere on the route (for example, the services in Scenario B2 take significantly less than 45s, which may or may not be planned), but positioning slack effectively (including the location of personnel or rolling stock reserves as well as the allocation of buffer time en route) can be challenging at the best of times [11]. Further issues can arise when the reason for the slack at a station is unclear - is it to absorb late running from earlier in the route, to allow for significant passenger flows, or both? The development of visualisation tools (Section 4.4) enabling operators to better understand the whole picture could help them to improve their strategies.

A range of rolling stock types are operated by the TOC, some of which are also used elsewhere on the UK network. Scenarios A1 and A2 were partly chosen in order to help gain insight in to the potential impact of different rolling stock types. It was difficult to make direct comparisons between different rolling stock types because most services are operated by largely homogeneous fleets: There is some variation in train length, but this is accounted for separately (Table 1). The services in scenario A1 originated on a different part of the network to those in A2, but services in both scenarios had the same destination and a comparable stopping pattern after Station A. There may be some confounding factors (for example, trains operating on one route may typically rely on more buffer time to catch up at Station A than those operating on others), but the longer dwell times observed for the diesel trains (A2) is consistent with the fact that the dispatch process can be longer for diesel trains than for electric trains. This is because diesel engines need to spool up before power reaches the wheels, resulting in additional (often significant) lag between the moment the driver engages power and the moment the wheels start to move. There are also other differences between the diesel and electric fleets here which would also help explain the increased dwell times in scenario A2. These include the width of the doors (the diesel fleet has narrower doors, which could reduce passenger throughput) and the respective age of the fleets: The diesel trains are older, with door mechanisms which may be less efficient.

3.4. Network-wide regression models

After analysing the initial scenarios, the whole dataset was used to build and test regression models using the same basic methodology, in order to provide input in to a network-wide visualisation tool (Section 4.4). The set of features was the same as those used for the regression modelling (Table 1).

To account for the diversity of station and service types, the data were first grouped as follows:

- Station
- Headcode group (the first two characters of the headcode)
- Direction (1 or 2, based on the way the routes were described in the Network Plan provided by the TOC)

There were 721 distinct station-headcode-direction groupings in the input data, and individual regression models - henceforth known as ’predictors’ - were built for each one. Not every predictor could be calibrated successfully, mainly due to the fact that a minimum sample size was chosen (in addition to a subset reserved for testing, a minimum of 100 data points were needed to calibrate each predictor). For the initial scenarios, care had been taken to maximise the homogeneity of the schedules in question, by manually specifying scheduled dwell times for the previous stops on the route. To have implemented this across the whole network would have entailed more complex groupings (adding scheduled dwell time at each of the two previous stations to the list above), increasing both the number of potential models and the likelihood of insufficient sample sizes. The trade-off is the risk that some of the outputs are adversely impacted by scheduling variations, but this was deemed acceptable at the 'proof of concept' stage.

The coverage of the network represented by these distinct station-headcode-direction groupings was not reviewed in detail at this stage, but it is known that some services were not covered (due to the use of a type of rolling-stock for which OTMR data were not provided). It was also discovered that the 'station' field in the dwell time data provided by the operator contained at least one entry which was not an actual station (in this case, it referred to scheduled pauses at a passing loop). At this initial stage of developing a proof of concept, the network data were not filtered to exclude such non-station stops, but they did not impact the scenarios considered in detail.

As with the regression modelling for the initial scenarios, input data used for calibration were filtered to exclude cases where the (absolute) difference between the actual and the scheduled dwell time exceeded 180 seconds. In contrast to the initial scenarios, the data were not filtered on arrival time. However, cases where the dwell time exceeded 600 seconds (10 minutes) were excluded, as longer dwell times were not felt to be representative of normal operations on the network or important when considering sub-threshold delays.

To provide outputs for the visualisation tool, a second set of regression models was also calibrated for the same dataset, using actual dwell time, rather than variation from the scheduled dwell time as the dependent variable. The set of features and method of implementation remained unchanged.

Table 3: Station dwell time categorisations

Dwell Category	Description
0	Exact on-time arrival and departure
1	Early arrival with lower than scheduled dwell time
2	Early arrival with greater than scheduled dwell time
3	Late arrival with greater than scheduled dwell time
4	Late arrival with lower than scheduled dwell time

3.5. Dwell Time Categorisation

Regression models (3.2) are a potentially useful way of quantifying some of the factors which cause variations in dwell time and of understanding how perturbations may propagate. However, variations in dwell time alone are not necessarily linked with sub-threshold delays; for example, greater than scheduled dwell times may be as a result of early arrivals (arising from slack in the previous sectional running times) and not delayed departures. Furthermore, when making operational decisions, it is not generally essential to predict station dwell times to the nearest second. In many cases, it is desirable simply to be able to predict overall trends and to identify potentially problematic services.

Figure 1 shows a snapshot for all services through Station B on a single day in April 2017. Services which deviated from their scheduled dwell time fit in to one of four quadrants, which can be used to categorise station dwell times. These categories (including an additional one for services which arrived and departed exactly on schedule) are given in Table 3.

By definition, trains in Category 1 leave the station ahead of schedule. A previous study of trains in London observed a number of occurrences when trains were dispatched and departed 'unduly early' [5]. No data about the

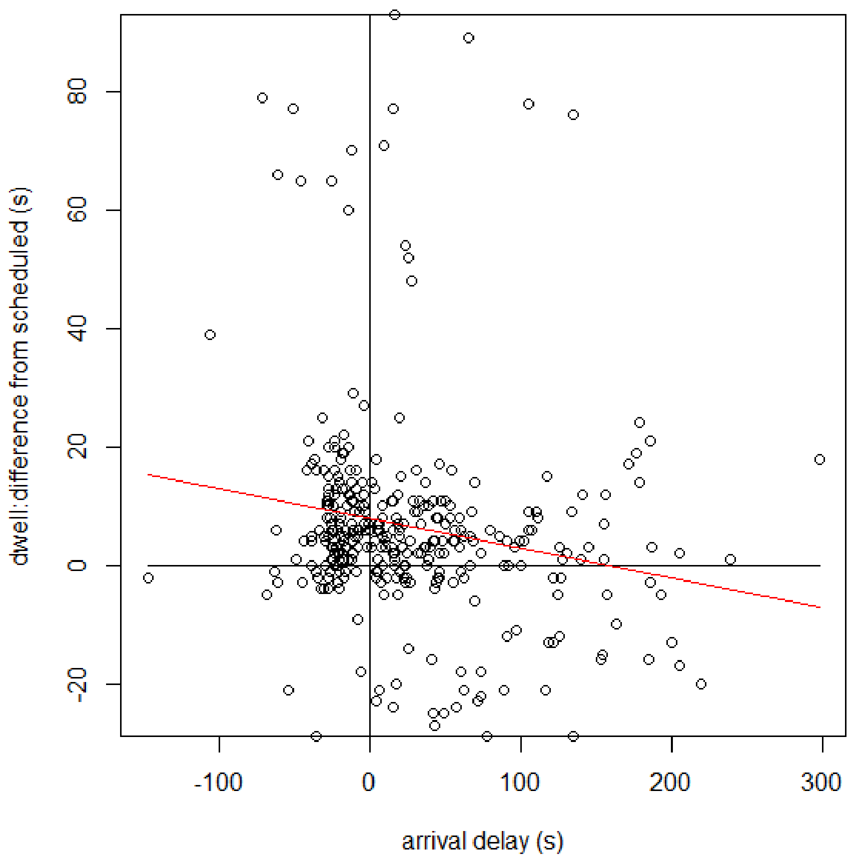

Figure 1: Difference from expected dwell time by arrival delay for Station B on a single day

dispatch process were available in this case, but trains were flagged as having departed 'unduly early' if wheelstart was 15 seconds or more before the public scheduled departure time. Public timetables are given to the minute, in contrast to the 30s granularity of the working timetable, which means that in many cases these services would have left more than 45s before the scheduled time in the working timetable. In some cases there is deliberate variation between the working timetable and the public timetable, allowing working times to be flexed for operational reasons whilst keeping the public timetable more consistent for marketing purposes.

Table 4 shows how the data from each of the initial scenarios breaks down

Table 4: Breakdown of the initial scenarios by dwell category

Scenario		A1	A2	B1	B2	C1	C2
Proportion of trains in each dwell category	0	0%	2%	1%	4%	3%	1%
	1	6%	18%	6%	3%	2%	0%
	2	7%	8%	1%	13%	28%	18%
	3	6%	4%	7%	70%	62%	77%
	4	81%	68%	85%	10%	6%	4%
Proportion of trains which left unduly early		0.3%	1.3%	2.6%	2.4%	1.0%	0.4%

in to each of the dwell categories and gives the proportion of services flagged as having departed unduly early.

3.6. Network-wide models to predict dwell classification

Being able to predict the different categories of station dwell (defined in Table 3) in real time across a network could enable operators and planners to identify potential problems and undesirable patterns. To achieve this aim, two Random Forest classifiers were developed. The first classifier predicted the classification of the dwell time according to Table 3. The second was a binary classifier, flagging up whether the dwell time was likely to contribute to a delay; for the purposes of this exercise a station stop was marked as contributing to a delay if (i) the train departed late and (ii) the dwell time was greater than scheduled.

Following the precedent set with the network-wide regression models (Section 3.4), data were grouped by station, headcode group and direction and individual predictors were built for each group (with the data filtered in the same way, and the same minimum sample size applied). The Random Forest classifiers were based on the same set of features used for the regression modelling (Table 1).

Table 5: Adjusted R^2 values for the chosen scenarios

Scenario	A1	A2	B1	B2	C1	C2
Adjusted R^2	0.45	0.41	0.67	0.37	0.43	0.62
Adjusted R^2 (peak trains only)	0.51	-	-	0.42	0.46	0.72
Adjusted R^2 (off-peak weekday only)	0.46	-	-	0.33	0.39	0.6
Adjusted R^2 (weekend trains only)	0.48	-	-	0.45	0.45	0.56

4. Results

4.1. Initial Regression Models

4.1.1. Adjusted R^2 values for each of the scenarios

Table 5 gives the adjusted R^2 values for the linear regression models which were generated for each of the scenarios. Where there were sufficient data, models were additionally fitted separately for peak trains, off-peak weekday trains and weekend trains.

4.1.2. Variation in R^2 values between peak, off-peak and weekend services

In line with the literature taken as a basis for the models [9] it was found that the adjusted R^2 was greatest for peak services, reflecting the fact that these services typically exhibited the least variation in station dwell (Table 2). Off-peak weekday services tended to exhibit more variation in station dwell than weekend services, and this is also reflected in the adjusted R^2 values.

Although the core off-peak service patterns are similar throughout the week, with Saturdays in particular mimicking the weekday off-peak pattern, there are several reasons why more variation is observed in station dwell times during the week. Firstly, those off-peak services which run immediately after peak time trains during the week (sometimes referred to as the 'shoulder peak') may still be subject to perturbations due to earlier peak time congestion. Indeed, some of the services which are categorised here as off-peak trains may have been classed as peak time trains earlier in their journey. Furthermore, shoulder peak services on a weekday can be subject to particularly high passenger flows as travellers choose to avoid paying much higher peak time fares when they can. The period covered by the data included some school holidays, which have more of an impact on weekday passenger flows than on weekend passenger flows, and the general mix of passengers is also likely to be different between off-peak weekday and weekend trains.

Weekend trains would be expected to be more dominated by leisure travellers, with much more business travel happening during the week. This will have some impact on passenger flows, and the potential for particularly crowded trains at certain times during the week. The passenger mix itself may also influence station dwell times, with leisure travellers being more likely to be encumbered (for example, by having luggage or travelling with children as well as being more likely to be elderly) and possibly less likely to be familiar with the system. These impacts cannot explicitly be inferred from the models but may be a reason for some of the unexplained variation.

Scenario C2 stands out as having a lower adjusted R^2 value for weekend services than for off-peak weekday services, despite the fact that the observed variation in dwell time is still lower at weekends. This could be explained by the fact that one of the preceding stations from which dwell time is an input to the regression model serves a sporting venue which hosts large events at weekends. It is likely that the station in question will have higher dwell times as crowds are dispersed immediately after an event, whilst Station C itself will not be impacted significantly by this crowding. Whereas crowding resulting from peak time travel patterns or serious disruption can build up along a route, the impact of a large weekend event can be more localised. In fact, other demand for the service may well drop as travellers know to avoid these times, and local people travelling home from an event may conceivably choose another route. This example highlights the need to appreciate the whole context when studying station dwell times, something which could be helped by the provision of visualisations.

4.1.3. An overview of the feature coefficients

The feature coefficients (for the model applied to all trains in the data) and associated p-values are given in Table 6 and Table 7 respectively. Some of the features listed in Table 1 were not found to be significant in any of the scenarios and hence are not included here; these factors include whether or not it was a weekend, whether or not there had been changes to the planned order of services, the dwell time of the previous service at the station and the dwell time of the same service the previous week. There are two main reasons for the insignificance of the boolean variable taking into account the impact of a weekend. The first is that peak time trains are by definition weekday services only, and so some of the weekday/weekend differences will instead be encapsulated by the boolean variable differentiating between peak and off-peak services. The second is that weekday services are subject to more variation than weekend services, with variation being as likely - if not more so - between individual weekdays than between weekdays as a whole and weekends (for example, more work trips take place on Tuesdays, Wednesdays and Thursdays than they do on Monday or Friday).

4.1.4. Further observations about peak and off-peak services

Table 7 shows that when considering all services together, scenario C1 is the only case in which the boolean to differentiate between peak and off-peak services is significant (although it should be noted that this is irrelevant for scenarios A2 and B1 which only had data for peak time services). The relatively large positive coefficient is consistent with the relatively large increase in dwell time observed for peak time services. The lack of impact of the peak on services in the other direction (scenario C2) can be explained by a number of factors: Passenger flows out of London are relatively low in the morning peak, whilst in the evening peak, crowding is predominantly due to people leaving London: In this case, an increase in passengers alighting at Station C is expected, but

Table 6: Regression model feature coefficients for the chosen scenarios

Scenario	A1	A2	B1	B2	C1	C2
previous_stop_dwell_sec	-	-0.24	0.37	0.2	0.11	0.62
second_prev_stop_dwell_sec	-0.07	-	0.06	0.03	0.29	-
previous_stop_dep_delay	-0.01	-	-	-	-	-
prev_service_same_line_dwell_sec	-	0.06	-	-	-	-
actual_length	-4.38	-2.67	-	-0.46	-	-1.19
previous_train_length	-	-	-4.96	-	-	-
peak_train_True	-	-	-	-	3.3	-

Table 7: Regression model feature p-values for the chosen scenarios

Scenario	A1	A2	B1	B2	C1	C2
previous_stop_dwell_sec	-	0	0	0	0	0
second_prev_stop_dwell_sec	0.00628	-	0	0.00002	0	-
previous_stop_dep_delay	0.00003	-	-	-	-	-
prev_service_same_line_dwell_sec	-	0.0053	-	-	-	-
actual_length	0	0	-	0	-	0
previous_train_length	-	-	0	-	-	-
peak_train_True	-	-	-	-	0.00042	-

large crowds of passengers on the platform waiting to board is not expected. For 'down' trains (scenario C2), Station C is closer to the origin of the service than it is for 'up' trains (scenario C1) and so there is less scope for peak time congestion on the network to have already perturbed the schedule.

4.1.5. The impact of disruption to the timetable

The fact that changes to the planned order of services is not significant for any of the scenarios may partly be due to the decision to exclude major delays (services arriving 20 minutes or more late at the station) as this would have minimised the likelihood of displacement of staff and other factors. Any impact of disruption across the network on passenger numbers would also have been captured by some of the other variables (such as the dwell at the previous stop). It should also be noted that the chosen stations have multiple platforms, which is likely to reduce the risk of platform crowding and congestion in the event that the order of services is disrupted; similarly, previous departures from the station may have left from other platforms, possibly explaining why they have no significant impact here.

4.1.6. The propagation of dwell delays along a route

For stations B and C, the coefficients for actual dwell time at the previous stations are positive (Table 6), which is consistent with the hypotheses that increased dwell times earlier in the route can be indicative of factors such as higher passenger loadings or a mechanical issue with the doors which would also be expected to increase the dwell time at the station in question. This positive correlation also demonstrates how dwell delays can propagate along a route, noting that as the variables are in terms of actual dwell time and not variation from scheduled dwell time, they can have a significant impact. Recasting the models in terms of variation from scheduled dwell time could have presented a confused picture here, given that (for example) scenario B1 appears to have a planned dwell which is below the scheduled dwell. Obtaining planned dwell timings (as opposed to scheduled dwell timings) from the operator for each service and station would allow enhancements to be made the model to better quantify how delays can propogate.

It is notable that the dwell time at the second previous station is insignificant for scenario C2 (Table 7), which is likely to be linked with the fact that the scheduled dwell for the second previous station is 60s, not 30s and therefore contains more buffer time to mask some of these impacts. For Station A, the coefficients for the variables relating to performance at the previous stations on the route are all negative, which is consistent with the having a large amount of buffer time to make up for earlier delays: If a train loses time at a station further down the line, this will be taken off the buffer time at Station A and hence reduce the observed dwell at Station A. The fact that in scenario A1 it is the delay departing the previous stop, not the dwell time, which is significant supports this theory.

4.1.7. The impact of train length

Where train length is significant, the coefficient is negative: Longer trains have reduced dwell times. One reason for this may be the positioning of the train on the platform [9], although the idea that the position of the trains is more 'rigid' when they are longer does not apply here (stop boards are used to ensure fixed stopping points for all types and length of rolling stock). The impact instead is likely to be because longer trains cover more of the platform, increasing the likelihood that passengers, especially unfamiliar travellers, are close to a door when the train arrives. Another reason that longer trains reduce dwell time is that fewer passengers board and alight at each door. Regular passengers know to spread themselves out along the platform, whilst dispatch staff can be good at encouraging less familiar passengers to spread out. This has been informally observed at Station A, and others like it, where both the guard and station staff are often proactive in ensuring that passengers use other available doors instead of queuing on the platform to wait for someone with luggage to board in front of them. The positive impact of the previous train length in scenario B1 (which covers peak time trains only) may be to do with variation in passenger flow during the peak period: For example, trains just before and at the beginning of the peak period may typically be shorter and less busy than those later in the peak. Further work would need to be done to investigate this.

4.2. Use of the regression models to make predictions across the whole network

4.2.1. Adjusted R^2 values

When models were applied to the whole dataset (Section 3.4), there was significant variation in the adjusted R^2 values for the different station-headcode-direction groupings. Thirteen station-headcode-direction groupings had an adjusted R^2 value which was greater than 0.9. All of them were valid stations, but majority of these had very small sample sizes, so it would be inappropriate to draw too many conclusions without obtaining a larger dataset and re-running the models. However, one station-headcode-grouping had a sample size of 2,465 and an adjusted R^2 of 0.94, providing a great deal of confidence that the model could explain most of the observed variation in station dwell time. Assessment of the station and services in question suggests a number of reasons for this:

- It is a provincial station and although it is served by different services from more than one TOC, there is limited scope for external factors (such as a major event in the vicinity) to lead to a significant build-up of crowds

- Most services in the headcode group are operated by relatively short (two carriage) trains, which is in keeping with the provincial nature of the route. This limits the scope for issues caused by uneven passenger distribution, especially given that the weather protection on the platform is ample for such short trains

- The particular headcode group in question has an homogenous stopping pattern throughout the day, meaning that there is no scope for variation in service patterns not accounted for by the model to have an impact

- Observed dwell delays tend to be confined to specific trains each day, consistent with when passenger flows are generally known to be higher along the whole route

- The station only has one platform in each direction, further limiting the possible impact of external factors not considered by the model (such as variation in facilities and weather protection between platforms, or the impact on passenger flow of a very late platform change)

- The nature of the route means that there is limited need for buffer in the schedule at this station to make up for lost time; there are no junctions or other external reasons for services to be perturbed immediately prior to this station stop

At the other end of the scale, some predictors had adjusted R^2 values approaching zero. Small sample sizes were possibly a factor in some cases, but the main reason for very low adjusted R^2 values is the impact of external factors which the models do not consider. The models do not include weather and environmental data (despite plans to the contrary noted in Section 3.1) or data about the degree of weather protection available at each station. An exception to this is the fact that some universal impacts (e.g. the possibility that boarding and alighting might take longer if everyone is wearing bulky clothes and coats) will be taken into account by the use of values of dwell time at previous stations. There are also passenger behavioural impacts which aren't taken into account, such as the 'late runners' who arrive after the bulk of passengers have boarded ([4]) or those who for whatever reason choose to hold the doors open and delay the dispatch process ([18]). Although dwell time at previous stations can indicate a potential impact of large passenger flows, some of the effects may not be captured at earlier stations en route. These include the fact that the increase in marginal delay with respect to the number of standees is non-linear ([21]) and there is a threshold above which the size of the step gap between the platform and the train slows down passenger flow ([4]). As evidenced by Scenario C2, external factors such as large events in the vicinity can also influence passenger numbers (Section 4.1.2).

The interaction between different services is also not taken into account in the model. Possible impacts include an increase in passengers waiting on the platform following the arrival of a connecting service, or a decrease in available buffer time due to arrival delays caused by unfavourable signalling at a junction. The 50 predictors with the lowest adjusted R^2 values all related to actual stations (as opposed to the passing loop in the data). However, some of them related to passing stations on single track sections of the network where waiting for trains in the other direction is a potentially dominant factor in dwell time variation not considered by the models.

4.2.2. Relative feature importance

When the individual predictive models were calibrated, the relative importance of the input features was automatically calculated by considering their impact on the adjusted R^2 value, and it was found that no single feature was consistently the most important across the network. For example, for 24% of station-headcode-direction groupings, the dwell time at the previous stop was found to be the most influential feature, whilst for another 19% train length came out on top.

4.2.3. Application of the regression models to separate test data

Although the R^2 value is a useful measure of the performance of a model, a better idea of the ability to predict dwell time variation was found by applying the predictors to the separate test sample (20% of the data) for the whole network, and the outputs are shown graphically in Figure 2. Although the percentage variation seems to be quite high, it is important to note that the timescales are very small, and the resolution of the data are comparatively low (i.e. the model is predicting single second variations based on data with a resolution of a single second). If the actual dwell time exceeds the scheduled dwell time by a second, a prediction of two seconds is a 100% difference. It is also noted from Figure 2 that the distribution of results is negatively skewed; indeed, the median difference between actual dwell variation and predicted dwell variation is -26%. The fact that the model overall tends to under-predict dwell time needs to be borne in mind and should be further investigated.

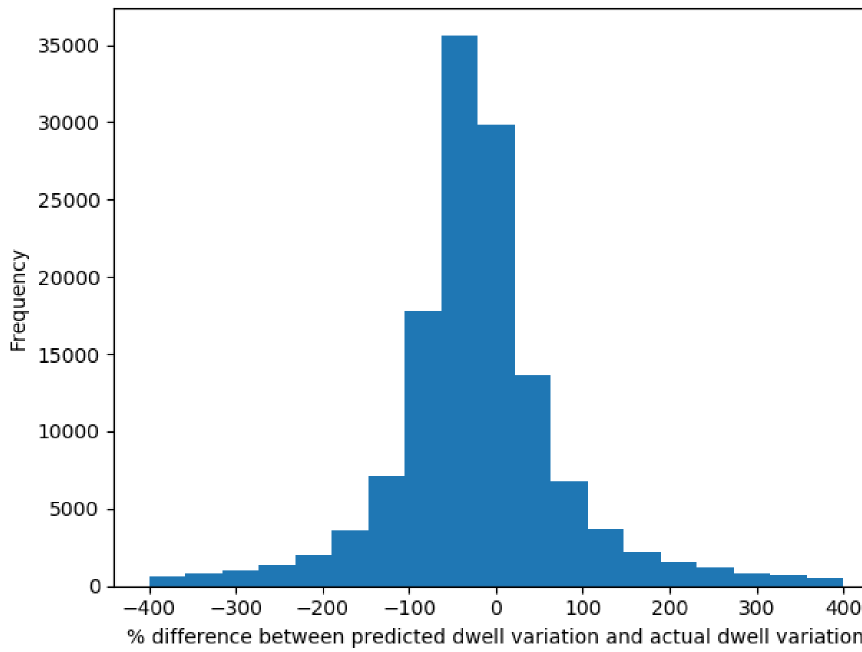

Figure 2: Variation between actual and predicted values across the test sample

4.3. Prediction of dwell time classification using Random Forests

4.3.1. Classifying station dwells by type

The dwell-time classifier models were applied to a separate test data set comprising 140,873 station dwells. The first model classified the station dwells according to Table 3, and the results are given in Table 8.

Table 8: Actual and predicted dwell classifications for test data across the whole network

Dwell Classification	Actual proportion in test data set (%)	Predicted proportion in test data set (%)
0	2	2
1	6	5
2	18	17
3	41	49
4	32	27

Although the predicted proportions seem to match the actual proportions favourably, the figures in Table 8 make no allowances for mis-classifications. Figure 3 illustrates the range of reported accuracies across all the individual predictive models, and it is clear that there are some station-headcode-direction groupings where the accuracy is very poor. As with the regression models, these classifiers may suffer from a lack of contextual information (including weather and platform crowding levels) and further work should be undertaken to review the groups which perform most poorly and investigate how the model can be improved.

4.3.2. Binary Classification

The results for the binary classifier, which flags instances where the dwell time is likely to contribute to a delay to the service, were much more positive (Figure 4). Analysis was undertaken, looking at how the data varied for different stations and services. Whereas certain stations have a high rate of instances where the dwell time exceeds the scheduled dwell time and contributes to a delay, the data aggregated by headcode group are more neutral. It is therefore suggested that investigating contextual factors at different stations, rather than for different services should be the priority.

It is crucial to understand not just how accurate the system is across the

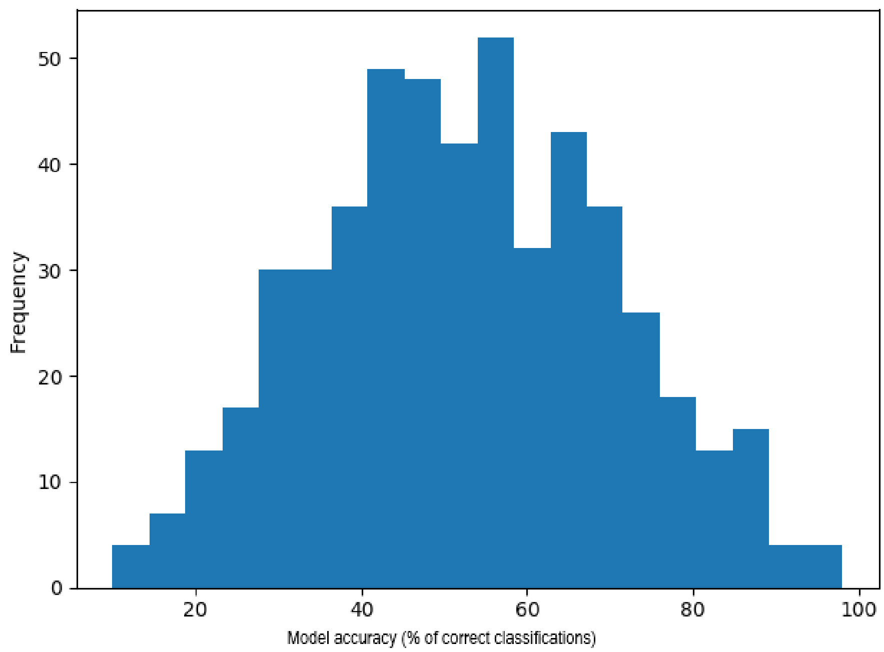

Figure 3: Frequency plot of the accuracy of the individual classifiers

network, but also the propensity for the system to show false positives or false negatives. If there is a tendency for the system to mainly show false positives, for example, there is a risk that operators may start to ignore the alerts. A review of the false positive and false negative rates showed significant variation between stations, with some stations recording very high levels of false positives and others very high levels of false negatives. This may be linked to the fact that planned dwell times can differ from timetabled dwell times at certain stations, and the actual allocation of buffer time is unclear from the data. In any case, the initial outputs from the research can help the operator prioritise the stations most in need of further investigation.

4.4. Visualisation tools

A web-mapping interface for visualisation was developed by the project team, building upon previous work developed in collaboration with RSSB [16], based

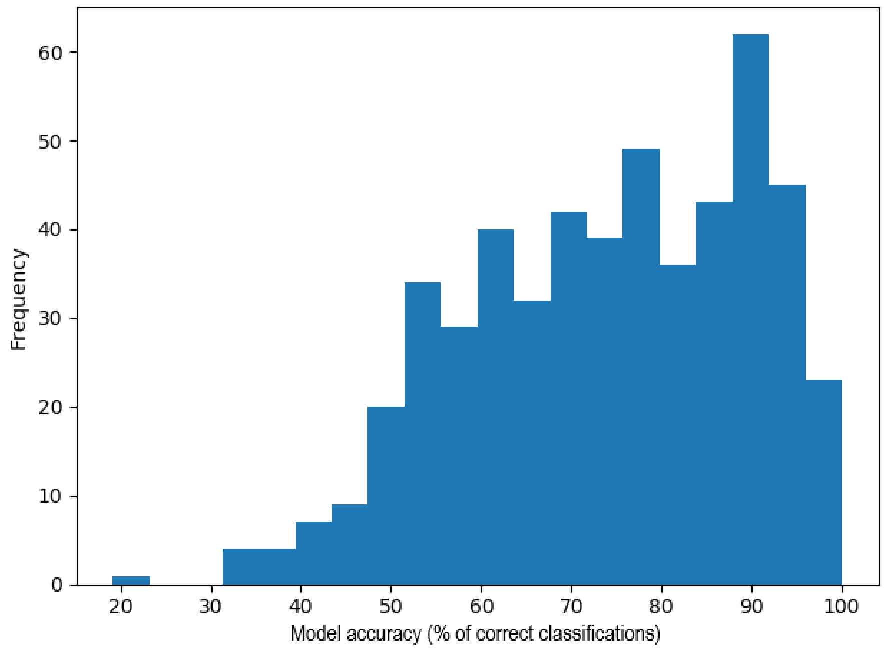

Figure 4: Frequency plot of the accuracy of the individual binary classifiers

on relevant open web standards and open source technologies and tools. The application, illustrated in Figure 5, presents an attractive, interactive map interface, allowing the user to explore, filter and visualise the dwell time data overlaid on a UK map showing the Rail Network. The image shown demonstrates the map interface used in 'DEMO MODE' where dwell values have been randomised in order to restrict access to the TOC data, where necessary.

A data selection control allows a user to filter displayed data by a range of dates; all days, weekdays (optionally including peak times only) or weekends; exclude dwell times greater than a certain number of seconds (to remove obvious outliers or extreme values); select one of the four dwell time/delay quadrants (mentioned above in Section 3.5); and choose an aggregation method for visualisation (mean, median, max, min).

Filtered values are displayed as coloured, scaled circles (green for negative dwell difference to red for positive dwell difference), with an actively updated legend to interpret those values shown. Clicking a circle on the map reveals more information about the dwell time data aggregated at that station, with an option to show a detailed breakdown by headcode, with histogram illustration of value ranges and frequencies.

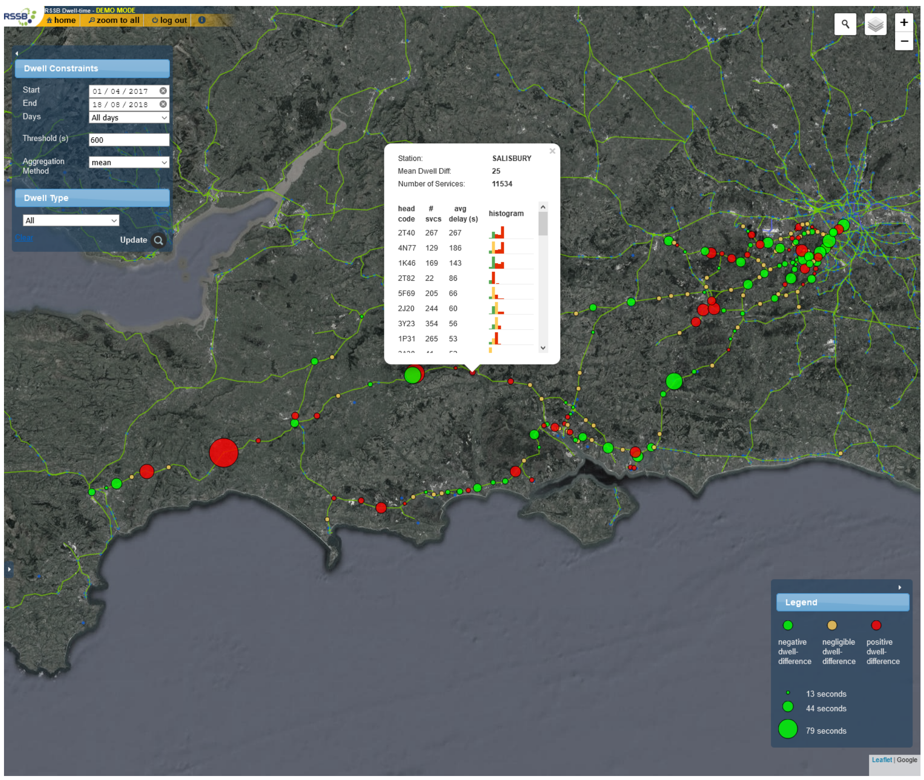

Figure 5: Web-mapping interface for interactive data visualisation

The web mapping application provides an effective tool to illustrate the spatial and temporal context of the dwell data, with potential to integrate other geographic locations and additional rail data sources.

Though live OTMR data feeds were not available at the time of development, the tool includes an Alerts module for displaying real-time dwell data (implemented using a simulated feed), illustrated in Figure 6. When invoked, the Alerts tool allows the user to choose from a list of Model Scenarios (a predefined list, but retrieved dynamically from the live database), then simulate the running of the predictive model in near-real time, animating through the model run period while displaying the model predictions as symbols on the map. The model run can be controlled with play/pause/stop controls and map symbols (coloured and sized in proportion to exceedance of predicted dwell values over a chosen threshold) can be clicked to reveal more information about the predicted dwell exceedance, actual dwell values (for previous stops), with potential in future to also include confidence in modelled values and recommended mitigations for a delay.

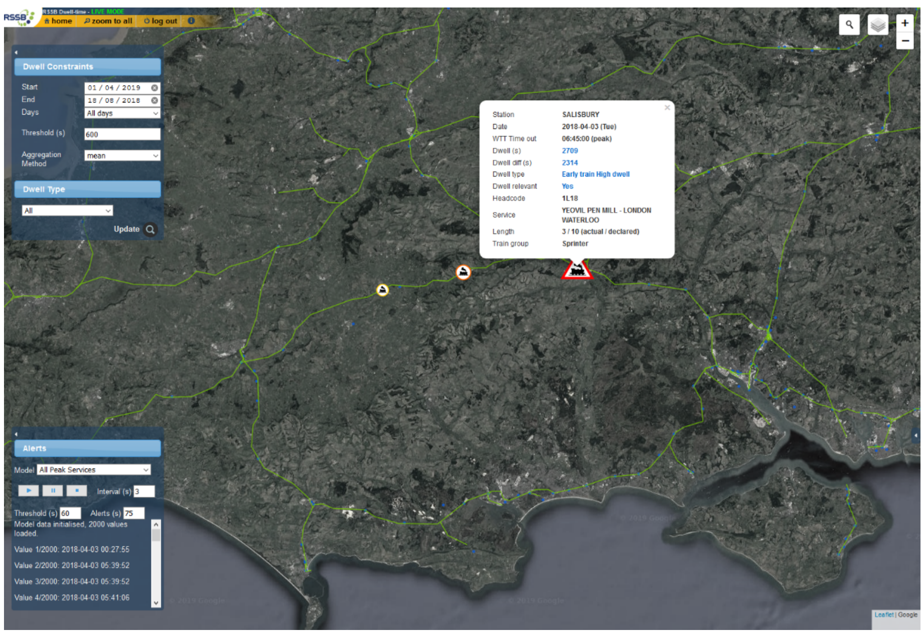

Figure 6: Alerts module for displaying real-time dwell data

An additional data analysis and visualisation application was developed using the open source "Shiny" R package, which allows a custom visual web interface to be created to the R system for statistical computing. The application includes a selection of useful tools for analysing the available dwell data and visualising the journey data through dynamically created graphs and charts. It also includes a more sophisticated (though experimental) Bayesian network model. The tool includes a set of named tabs, with the following functions.

Scheduled vs Actual Departure Times For a selected date, journey and time, a graph is displayed showing both scheduled and actual departure times, showing at a glance if a train ran to schedule (Figure 7 - top). A helpful variation of this graph, for future development, would be to display a comparision of departure time lateness against arrival time lateness, enabling users to identify which services are losing time.

Dwell at a Station over a Day For a selected date and station, a graph is shown of the dwell (difference from scheduled, so a negative time means the dwell was less than scheduled) for all trains. The actual departure time is given on the x-axis so it is possible to see patterns of dwell changing throughout the day (Figure 7 - middle).

Dwell time vs Arrival Delay For a selected date and station, a graph is shown which shows the arrival delay on the x-axis and the difference in dwell from that scheduled. This is helpful in demonstrating that longer and shorter dwell times can be for different reasons (Figure 7 - bottom).

Monthly Mean Dwell per Hour For a given station and year the mean dwell difference from schedule is calculated for each month and for each hour of the day, allowing users to identify times or periods with consistently greater than expected dwell times.

Stations With Mean Dwell > 20s above Scheduled A series of graphs are created for stations where the mean dwell of all trains through the station over the time period is more than 20 seconds above the scheduled dwell - indicating consistently high dwell. For each station there are two graphs. A graph of the arrival time is displayed, which helps to indicate if the dwell might just be because a train is arriving consistently early or if it might be for some other reason. The second graph is of the dwell times. Red lines on each graph show the scheduled dwell or arrival time.

Bayesian Modelling A proof of concept model was created linking all stations on a route in a Bayesian network, to enable users to dynamically investigate the propagation of delays through the rail network. A leaving delay can be set for a specific station and the predicted effect on arrival delay at subsequent stations is displayed, demonstrating how the effects of the delay propagate through the system. There is great potential for further development and more robust testing of this tool, which would be a focus for future work.

5. Conclusions

Small fluctuations in station dwell time can have an impact on overall service performance, but the scale of the impact and the possible causes of the fluctuations are not always widely understood or well communicated. This is partly because a number of different factors can influence station dwell time, many of which (such as detailed passenger flows) can be hard to monitor on a large scale. This paper built on previous research to investigate whether train operation data can give some insight into dwell time fluctuations and whether it can be used to provide sufficient information for railway operators to understand where problems are likely to occur.

Models were developed to predict variation from scheduled dwell time, categorise the predicted variation and to flag where a station dwell time might contribute to a delay. Encouraging results were obtained, demonstrating that such models can help make predictions about station dwell times and can explain some of the issues. Where services have limited exposure to external factors, it was shown that particularly good model fits could be achieved.

Predicting the fluctuations alone can be of limited use because it is not possible to tell whether the train arrived early or left late, although the regression models applied to a set of specific scenarios provided some insight into the factors which lead to variation. In the case of the London stations, it was clear that dwell delays can propagate along a route, reinforcing the need to pay attention to initial sub-threshold delays. Categorisation of the station dwell times

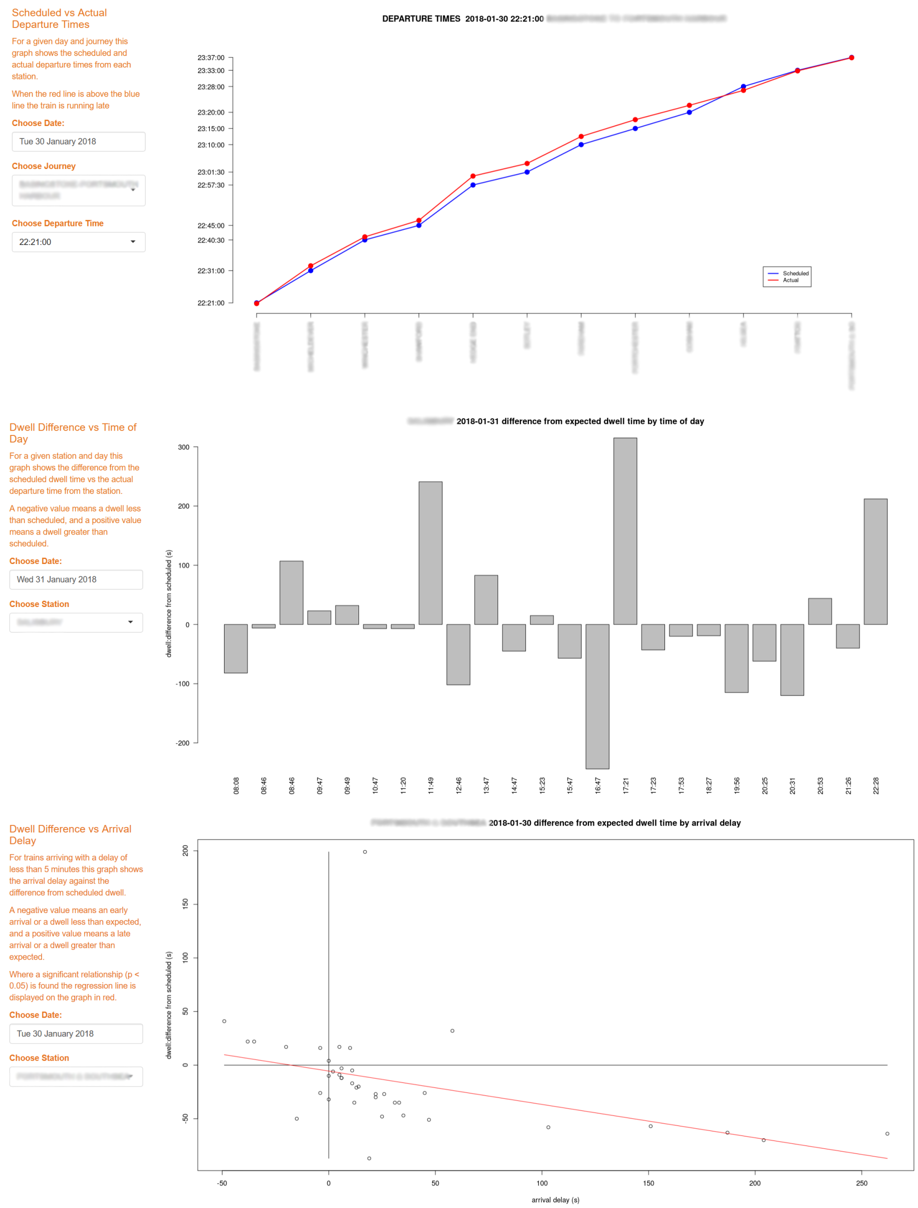

Figure 7: R Shiny data analysis and visualisation interface (station names have been blurred to restrict access to TOC data)

provided a way of identifying where sub-threshold variations might contribute to delays, and could help operators prioritise stations of interest.

Analysis of the scenarios showed the impact of buffer time (or 'slack') and highlighted the practice of allowing planned dwell times to differ from scheduled dwell times, in order to maintain 45s dwells from a timetabling system devised to the nearest 30s. This can mask sub-threshold delays; an adverse effect of slack in the timetable is a reduction in pressure to keep to time, and a lack of visibility of sub-threshold impacts may only serve to exacerbate this problem. This research has shown the importance of considering the system as a whole and demonstrated how visualisations could be provided to help operators do that. The importance of contextual factors has also been shown, and visualising the network may help operators identify and apply their own contextual knowledge (such as the presence of a major events venue).

If data could be obtained from the operator giving planned dwell times for each service and station (not just generic scheduled dwells), the outputs of the models would be enhanced and they could be really used to show how delays propagate through the system. There are a number of additional ways in which the models could be enhanced. Firstly, more details about the rolling stock itself (as well as train length) could be included. Although there are some caveats when it comes to including train type and train length data, the data do exist in some form. In theory, adding train type to the models is trivial; in practice, there are only a few services where the stock type varies, and these would need to be identified. Secondly, contextual factors including weather and passenger numbers, should be considered. It is understood that work is ongoing with at least one train operator to monitor passenger numbers, and if this data could be obtained it would likely make a positive difference to model accuracy. It would also help to discern exactly why some of the known factors, such as train length, make a difference. Finally, it would be good to consider possible interactions between different features.

Even at this early stage, without such enhancements, the models have shown great potential to provide important insights for railway operators. The outputs were used to support prototype visualisation tools which can help operators understand where dwell time delays cause particular problems on the network, and to make informed decisions about potential mitigation actions. These visualisation tools have been designed to work with real-time data feeds, paving the way for a real-time alert system, and demonstrating the potential for Bayesian Modelling to simulate the effect of delay propagation within the network. For busy and congested networks, where operators are under increasing pressure to reduce dwell time delays, such systems could have a significant impact. The model outputs can also be used to identify particularly problematic stations and services; even cases where the model fit is poor can be used to prioritise stations for further investigation and may lead to new insights about station characteristics which help or hinder train dwells.

6. Acknowledgements

This research was funded by RSSB as part of its Data Sandbox: Improving Network Performance competition. It was supported by rail industry partners, including South Western Railway, who provided data and operational insights and Network Rail who provided access to their Network Model.

References

[1] Carey, M., 1998. Optimizing scheduled times, allowing for behavioural response. Transportation Research Part B: Methodological 32B, 329–342. doi:10.1016/S0191-2615(97)00039-8.

[2] D'Acierno, L., Botte, M., Placido, A., Caropreso, C., Montella, B., 2017. Methodology for Determining Dwell Times Consistent with Passenger Flows in the Case of Metro Services. Urban Rail Transit 3, 73–89. doi:10.1007/s40864-017-0062-4.

[3] Goverde, R.M.P., Hansen, I.A., 2001. Delay Propagation and Process Management at Railway Stations. Proceedings CD-ROM of the World Conference on Railway Research (WCRR 2001) , 1–10.

[4] Harris, N., Graham, D., Anderson, R., Li, H., 2014. The Impact of Urban Rail Boarding and Alighting Factors. Transportation Research Board 93rd Annual Meeting , 1–13.

[5] Harris, N.G., 2015. A European Comparison of Station Stop Delays. International Congress on Advanced Railway Engineering , 1–6.

[6] Jiang, Z., Xie, C., Ji, T., Zou, X., 2015. Dwell time modelling and optimized simulations for crowded rail transit lines based on train capacity. Promet - Traffic - Traffico 27, 125–135. doi:10.7307/ptt.v27i2.1487.

[7] Jong, J.C., Chang, E.F., 2011. Investigation and estimation of train dwell time for timetable planning, in: Proceedings of 9th World Congress on Railway Research.

[8] KFH Group, Kittelson & Assoc Inc, Parsons Brinckerhoff Inc, Texas A&M Transportation Institute, Arup, 2013. Transit Capacity and Quality of Service Manual, Third Edition. Technical Report. Washington, D.C. URL: https://www.nap.edu/catalog/24766, doi:10.17226/24766.

[9] Li, D., Daamen, W., Goverde, R.M.P., 2016. Estimation of train dwell time at short stops based on track occupation event data: A study at a Dutch railway station. Journal of Advanced Transportation 50, 877–896. URL: http://doi.wiley.com/10.1002/atr.1380, doi:10.1002/atr.1380.

[10] Moncrieff, K., 2015. Designing passenger information for dwell time to support Thameslink high capacity infrastructure. The Fifth International Rail Human Factors Conference , 9URL: https://rhf2015.exordo.com/files/papers/64/final_draft/064_.pdf.

[11] Olsson, N.O., Haugland, H., 2004. Influencing factors on train punctuality - Results from some Norwegian studies. Transport Policy 11, 387–397. doi:10.1016/j.tranpol.2004.07.001.

[12] Open Rail Data Wiki, 2019. Main page — open rail data wiki,. URL: https://wiki.openraildata.com//index.php?title=Main_Page&oldid=2425. [Online; accessed 14-October-2020].

[13] Pritchard, J., 2019. Estimating the number of people in a train carriage: using data to provide future forecasts and current estimates in real time ., in: Proceedings of the World Congress on Railway Research, pp. 1–6. URL: https://www.sparkrail.org/Lists/Records/DispForm.aspx?ID=26689.

[14] Puong, A., 2000. Dwell time model and analysis for the MBTA red line. Massachusetts Institute of Technology Research Memo , 1–10.

[15] RSSB, 2017. How can train dwell times be reduced? Technical Report July. RSSB. URL: https://www.sparkrail.org/_layouts/15/Rssb.Spark/Attachments.ashx?Id=75NEMTS3ZVHP-8-12926.

[16] Sadler, J., Griffin, D., Gilchrist, A., Austin, J., Kit, O., Heavisides, J., 2016. GeoSRM - online geospatial safety risk model for the GB rail network. IET Intelligent Transport Systems 10, 17–24. doi:10.1049/iet-its.2015.0038.

[17] San, H.P., Masirin, M.I.M., 2016. Train dwell time models for rail passenger service. MATEC Web of Conferences 47. URL: https://www.matec-conferences.org/articles/matecconf/abs/2016/10/matecconf_iconcees2016_03005/matecconf_iconcees2016_03005.html, doi:10.1051/matecconf/20164703005.

[18] Volovski, M., Ieronymaki, E.S., Cao, C., O'Loughlin, J.P., 2020. Subway station dwell time prediction and user-induced delay. Transportmetrica A: Transport Science , 1–28URL: https://doi.org/10.1080/23249935.2020.1798555, doi:10.1080/23249935.2020.1798555.

[19] Watts, D., Jackson, K., Clarke, T., Williamson, T., Holloway, C., 2015. Passenger boarding time, mobility and dwell time for High Speed 2, in: The

Fifth International Rail Human Factors Conference, p. 11. URL: https://rhf2015.exordo.com/files/papers/74/final_draft/074.pdf.

[20] Weston, J.G., 1989. Train service model-technical guide. London underground operational research note 89, 18.

[21] Yuan, J., Medeossi, G., 2014. Statistical analysis of train delays and movements. Railway Timetabling & Operations; Hansen, I., Pachl, J., Eds , 217–236.