NETWORK IMPACTS OF A ROAD CAPACITY REDUCTION: EMPIRICAL ANALYSIS & MODEL PREDICTIONS

David Watling[1], David Milne[1] & Stephen Clark[2]

Abstract – In spite of their widespread use in policy design and evaluation, relatively little evidence has been reported on how well traffic equilibrium models predict real network impacts. Here we present what we believe to be the first paper that together analyses the explicit impacts on observed route choice of an actual network intervention and compares this with the before-and-after predictions of a network equilibrium model. The analysis is based on the findings of an empirical study of the travel time and route choice impacts of a road capacity reduction. Time-stamped, partial licence plates were recorded across a series of locations, over a period of days both with and without the capacity reduction, and the data were 'matched' between locations using special-purpose statistical methods. Hypothesis tests were used to identify statistically significant changes in travel times and route choice, between the periods of days with and without the capacity reduction. A traffic network equilibrium model was then independently applied to the same scenarios, and its predictions compared with the empirical findings. From a comparison of route choice patterns, a particularly influential spatial effect was revealed of the parameter specifying the relative values of distance and travel time assumed in the generalised cost equations. When this parameter was 'fitted' to the data without the capacity reduction, the network model broadly predicted the route choice impacts of the capacity reduction, but with other values it was seen to perform poorly. The paper concludes by discussing the wider practical and research implications of the study's findings.

1. INTRODUCTION

It is well known that altering the localised characteristics of a road network, such as a planned change in road capacity, will tend to have both direct and indirect effects. The direct effects are imparted on the road itself, in terms of how it can deal with a given demand flow entering the link, with an impact on travel times to traverse the link at a given demand flow level. The indirect effects arise due to drivers changing their travel decisions, such as choice of route, in response to the altered travel times. There are many practical circumstances in which it is desirable to forecast these direct and indirect impacts in the context of a systematic change in road capacity.

For example, in the case of proposed road widening or junction improvements, there is typically a need to justify economically the required investment in terms of the benefits that will likely accrue. There are also several examples in which it is relevant to examine the impacts of road capacity reduction. For example, if one proposes to reallocate road space between alternative modes, such as increased bus and cycle lane provision or a pedestrianisation scheme, then typically a range of alternative designs exist which may differ in their ability to accommodate efficiently the new traffic and routing patterns. Through mathematical modelling, the alternative designs may be tested in a simulated environment and the most efficient selected for implementation. Even after a particular design is selected, mathematical models may be used to adjust signal timings to optimise the use of the transport system. Road capacity may also be affected periodically by maintenance to essential services (e.g. water, electricity) or the road itself, and often this can lead to restricted access over a period of days and weeks. In such cases, planning authorities may use modelling to devise suitable diversionary advice for drivers, and to plan any temporary changes to traffic signals or priorities. Berdica (2002) and Taylor et al (2006) suggest more of a pro-active approach, proposing that

---

models should be used to test networks for potential vulnerability, before any reduction materialises, identifying links which if reduced in capacity over an extended period[3] would have a substantial impact on system performance.

There are therefore practical requirements for a suitable network model of travel time and route choice impacts of capacity changes. The dominant method that has emerged for this purpose over the last decades is clearly the network equilibrium approach, as proposed by Beckmann et al (1956) and developed in several directions since. The basis of using this approach is the proposition of what are believed to be 'rational' models of behaviour and other system components (eg link performance functions), with site-specific data used to tailor such models to particular case studies. Cross-sectional forecasts of network performance at specific road capacity states may then be made, such that at the time of any 'snapshot' forecast, drivers' route choices are in some kind of individually-optimum state. In this state, drivers cannot improve their route selection by a unilateral change of route, at the snapshot travel time levels.

The accepted practice is to 'validate' such models on a case-by-case basis, by ensuring that the model—when supplied with a particular set of parameters, input network data and input origin-destination demand data—reproduces current measured mean link traffic flows and mean journey times, on a sample of links, to some degree of accuracy (see, for example, the practical guidelines in TMIP, 1997, and Highways Agency, 2002). This kind of aggregate level, cross-sectional validation to existing conditions persists across a range of network modelling paradigms, ranging from static and dynamic equilibrium (Florian & Nguyen, 1976; Leonard & Tough, 1979; Stephenson & Teply, 1984; Matzoros et al, 1987; Janson et al, 1986; Janson, 1991) to micro-simulation approaches (Laird et al, 1999; Ben-Akiva et al, 2000; Keenan, 2005).

While such an approach is plausible, it leaves many questions unanswered, and we would particularly highlight two:

1. The process of calibration and validation of a network equilibrium model may typically occur in a cycle. That is to say, having initially calibrated a model using the base data sources, if the subsequent validation reveals substantial discrepancies in some part of the network, it is then natural to adjust the model parameters (including perhaps even the OD matrix elements) until the model outputs better reflect the validation data[4]. In this process, then, we allow the adjustment of potentially a large number of network parameters and input data in order to replicate the validation data, yet these data themselves are highly aggregate, existing only at the link level. To be clear here, we are talking about a level of coarseness even greater than that in aggregate choice models, since we cannot even infer from link-level data the aggregate shares on alternative routes or OD movements. The question that arises is then: how many different combinations of parameters and input data values might lead to a similar link-level validation, and even if we knew the answer to this question, how might we choose between these alternative combinations? In practice, this issue is typically neglected, meaning that the 'validation' is a rather weak test of the model.

2. Since the data are cross-sectional in time (i.e. the aim is to reproduce current base conditions in equilibrium), then in spite of the large efforts required in data collection, no empirical evidence is routinely collected regarding the model's main purpose, namely its ability to predict changes

---

[3]Clearly, more sporadic and less predictable reductions in capacity may also occur, such as in the case of breakdowns and accidents, and environmental factors such as severe weather, floods or landslides (see, for example, Iida, 1999), but the responses to such cases are outside the scope of the present paper.

[4]Some authors have suggested more systematic, bi-level type optimization processes for this fitting process (e.g. Xu et al, 2004), but this has no material effect on the essential points above.

in behaviour and network performance under changes to the network/demand. This issue is exacerbated by the aggregation concerns in point 1: the 'ambiguity' in choosing appropriate parameter values to satisfy the aggregate, link-level, base validation strengthens the need to independently verify that, with the selected parameter values, the model responds reliably to changes. Although such problems—of fitting equilibrium models to cross-sectional data—have long been recognised by practitioners and academics (see, e.g., Goodwin, 1998), the approach described above remains the state-of-practice.

Having identified these two problems, how might we go about addressing them? One approach to the first problem would be to return to the underlying formulation of the network model, and instead require a model definition that permits analysis by statistical inference techniques (see, for example, Nakayama et al, 2009). In this way, we may potentially exploit more information in the variability of the link-level data, with well-defined notions (such as maximum likelihood) allowing a systematic basis for selection between alternative parameter value combinations.

However, this approach is still using rather limited data and it is natural not just to question the model but also the data that we use to calibrate and validate it. Yet this is not altogether straightforward to resolve. As Mahmassani & Jou (2000) remarked: 'A major difficulty … is obtaining observations of actual trip-maker behaviour, at the desired level of richness, simultaneously with measurements of prevailing conditions'. For this reason, several authors have turned to simulated gaming environments and/or stated preference techniques to elicit information on drivers' route choice behaviour (e.g. Mahmassani & Herman, 1990; Iida et al, 1992; Khattak et al, 1993; Vaughn et al, 1995; Wardman et al, 1997; Jou, 2001; Chen et al, 2001). This provides potentially rich information for calibrating complex behavioural models, but has the obvious limitation that it is based on imagined rather than real route choice situations.

Aside from its common focus on hypothetical decision situations, this latter body of work also signifies a subtle change of emphasis in the treatment of the overall network calibration problem. Rather than viewing the network equilibrium calibration process as a whole, the focus is on particular components of the model; in the cases above, the focus is on that component concerned with how drivers make route decisions. If we are prepared to make such a component-wise analysis, then certainly there exists abundant empirical evidence in the literature, with a history across a number of decades of research into issues such as the factors affecting drivers' route choice (e.g. Wachs, 1967; Huchingson et al, 1977; Abu-Eiseh & Mannering, 1987; Antonisse et al, 1989; Bekhor et al, 2002; Liu et al, 2004), the nature of travel time variability (e.g. Smeed & Jeffcoate, 1971; Montgomery & May, 1987; May et al, 1989; McLeod et al, 1993), and the factors affecting traffic flow variability (Bonsall et al, 1984; Huff & Hanson, 1986; Ribeiro, 1994; Rakha & Van Aerde, 1995; Fox et al, 1998).

While these works provide useful evidence for the network equilibrium calibration problem, they do not provide a framework in which we can judge the overall 'fit' of a particular network model in the light of uncertainty, ambient variation and systematic changes in network attributes, be they related to the OD demand, the route choice process, travel times or the network data. Moreover, such data does nothing to address the second point made above, namely the question of how to validate the model forecasts under systematic changes to its inputs. The studies of Mannering et al (1994) and Emmerink et al (1996) are distinctive in this context in that they address some of the empirical concerns expressed in the context of travel information impacts, but their work stops at the stage of the empirical analysis, without a link being made to network prediction models. The focus of the present paper therefore is both to present the findings of an empirical study and to link this empirical evidence to network forecasting models.

More recently, Zhu et al (2010) analysed several sources of data for evidence of the traffic and behavioural impacts of the I-35W bridge collapse in Minneapolis. Most pertinent to the present

paper is their location-specific analysis of link flows at 24 locations; by computing the root mean square difference in flows between successive weeks, and comparing the trend for 2006 with that for 2007 (the latter with the bridge collapse), they observed an apparent transient impact of the bridge collapse. They also showed there was no statistically-significant evidence of a difference in the pattern of flows in the period September-November 2007 (a period starting six weeks after the bridge collapse), when compared with the corresponding period in 2006. They suggested that this was indicative of the length of a 're-equilibration process' in a conceptual sense, though did not explicitly compare their empirical findings with those of a network equilibrium model.

The structure of the remainder of the paper is as follows. In §2 we describe the process of selecting the real-life problem to analyse, together with the details and rationale behind the survey design. Following this, §3 describes the statistical techniques used to extract information on travel times and routing patterns from the survey data. Statistical inference is then considered in §4, with the aim of detecting statistically significant explanatory factors. In §5 comparisons are made between the observed network data and those predicted by a network equilibrium model. Finally, in §6 the conclusions of the study are highlighted, and recommendations made for both practice and future research.

2. EXPERIMENTAL DESIGN

The ultimate objective of the study was to compare actual data with the output of a traffic network equilibrium model, specifically in terms of how well the equilibrium model was able to correctly forecast the impact of a systematic change applied to the network. While a wealth of surveillance data on link flows and travel times is routinely collected by many local and national agencies, we did not believe that such data would be sufficiently informative for our purposes. The reason is that while such data can often be disaggregated down to small time step resolutions, the data remains aggregate in terms of what it informs about driver response, since it does not provide the opportunity to explicitly trace vehicles (even in aggregate form) across more than one location. This has the effect that observed differences in link flows might be attributed to many potential causes: it is especially difficult to separate out, say, ambient daily variation in the trip demand matrix from systematic changes in route choice, since both may give rise to similar impacts on observed link flow patterns across recorded sites. While methods do exist for reconstructing OD and network route patterns from observed link data (e.g. Yang et al, 1994), these are typically based on the premise of a valid network equilibrium model: in this case then, the data would not be able to give independent information on the validity of the network equilibrium approach.

For these reasons it was decided to design and implement a purpose-built survey. However, it would not be efficient to extensively monitor a network in order to wait for something to happen, and therefore we required advance notification of some planned intervention. For this reason we chose to study the impact of urban maintenance work affecting the roads, which UK local government authorities organise on an annual basis as part of their 'Local Transport Plan'. The city council of York, a historic city in the north of England, agreed to inform us of their plans and to assist in the subsequent data collection exercise.

Based on the interventions planned by York CC, the list of candidate studies was narrowed by considering factors such as its propensity to induce significant re-routing and its impact on the peak periods. Effectively the motivation here was to identify interventions that were likely to have a large impact on delays, since route choice impacts would then likely be more significant and more easily distinguished from ambient variability. This was notably at odds with the objectives of York CC, in that they wished to minimise disruption, and so where possible York CC planned interventions to take place at times of day and of the year where impacts were minimised; therefore

our own requirement greatly reduced the candidate set of studies to monitor. A further consideration in study selection was its timing in the year for scheduling before/after surveys so to avoid confounding effects of known significant 'seasonal' demand changes, e.g. the impact of the change between school semesters and holidays. A further consideration was York's role as a major tourist attraction, which is also known to have a seasonal trend. However, the impact on car traffic is relatively small due to the strong promotion of public transport and restrictions on car travel and parking in the historic centre. We felt that we further mitigated such impacts by subsequently choosing to survey in the morning peak, at a time before most tourist attractions are open.

Aside from the question of which intervention to survey was the issue of what data to collect. Within the resources of the project, we considered several options. We rejected stated preference survey methods as, although they provide a link to personal/socio-economic drivers, we wanted to compare actual behaviour with a network model; if the stated preference data conflicted with the network model, it would not be clear which we should question most. For revealed preference data, options considered included (i) self-completion diaries (Mahmassani & Jou, 2000), (ii) automatic tracking through GPS (Jan et al, 2000; Quiroga et al, 2000; Taylor et al, 2000), and (iii) licence plate surveys (Schaefer, 1988). Regarding self-completion surveys, from our own interview experiments with self-completion questionnaires it was evident that travellers find it relatively difficult to recall and describe complex choice options such as a route through an urban network, giving the potential for significant errors to be introduced. The automatic tracking option was believed to be the most attractive in this respect, in its potential to accurately map a given individual's journey, but the negative side would be the potential sample size, as we would need to purchase/hire and distribute the devices; even with a large budget, it is not straightforward to identify in advance the target users, nor to guarantee their cooperation.

Licence plate surveys, it was believed, offered the potential for compromise between sample size and data resolution: while we could not track routes to the same resolution as GPS, by judicious location of surveyors we had the opportunity to track vehicles across more than one location, thus providing route-like information. With time-stamped licence plates, the matched data would also provide journey time information. The negative side of this approach is the well-known potential for significant recording errors if large sample rates are required. Our aim was to avoid this by recording only partial licence plates, and employing statistical methods to remove the impact of 'spurious matches', i.e. where two different vehicles with the same partial licence plate occur at different locations.

Moreover, extensive simulation experiments (Watling & Maher, 1992; Watling, 1994) had previously shown that these latter statistical methods were effective in recovering the underlying movements and travel times, even if only a relatively small part of the licence plate were recorded, in spite of giving a large potential for spurious matching. We believed that such an approach reduced the opportunity for recorder error to such a level to suggest that a 100% sample rate of vehicles passing may be feasible. This was tested in a pilot study conducted by the project team, with dictaphones used to record a 100% sample of time-stamped, partial licence plates. Independent, duplicate observers were employed at the same location to compare error rates; the same study was also conducted with full licence plates. The study indicated that 100% surveys with dictaphones would be feasible in moderate traffic flow, but only if partial licence plate data were used in order to control observation errors; for higher flow rates or to obtain full number plate data, video surveys should be considered. Other important practical lessons learned from the pilot included the need for clarity in terms of vehicle types to survey (e.g. whether to include motor-cycles and taxis), and of the phonetic alphabet used by surveyors to avoid transcription ambiguities.

Based on the twin considerations above of planned interventions and survey approach, several candidate studies were identified. For a candidate study, detailed design issues involved identifying:

likely affected movements and alternative routes (using local knowledge of York CC, together with an existing network model of the city), in order to determine the number and location of survey sites; feasible viewpoints, based on site visits; the timing of surveys, e.g. visibility issues in the dark, winter evening peak period; the peak duration from automatic traffic flow data; and specific survey days, in view of public/school holidays. Our budget led us to survey the majority of licence plate sites manually (partial plates by audiotape or, in low flows, pen and paper), with video surveys limited to a small number of high-flow sites. From this combination of techniques, 100% sampling rate was feasible at each site. Surveys took place in the morning peak due both to visibility considerations and to minimise conflicts with tourist/special event traffic. From automatic traffic count data it was decided to survey the period 7:45–9:15 as the main morning peak period. This design process led to the identification of two studies:

Lendal Bridge study (Figure 1). Lendal Bridge, a critical part of York's inner ring road, was scheduled to be closed for maintenance from September 2000 for a duration of several weeks. To avoid school holidays, the 'before' surveys were scheduled for June and early September. It was decided to focus on investigating a significant southwest-to-northeast movement of traffic, the river providing a natural barrier which suggested surveying the six river crossing points (C, J, H, K, L, M in Fig 1). In total, thirteen locations were identified for survey, in an attempt to capture traffic on both sides of the river as well as a crossing.

Fishergate study (Figure 2). The partial closure (capacity reduction) of the street known as Fishergate, again part of York's inner ring road, was scheduled for July 2001 to allow repairs to a collapsed sewer. Survey locations were chosen in order to intercept clockwise movements around the inner ring road, this being the direction of the partial closure. A particular aim was to detect rerouting from Fulford Road (site E in Fig 2), the main radial affected, with F and K monitoring local diversion possibilities, and sites G, H, I, J to capture wider-area diversion.

In both studies, the plan was to survey the selected locations in the morning peak over a period of approximately twenty weekdays, covering the three periods before, during and after the intervention, with the days selected so as to avoid any known public holidays or special events.

In the Lendal Bridge study, while the 'before' surveys proceeded as planned, the bridge's actual first day of closure on September 11th 2000 also marked the beginning of the UK fuel protests (BBC, 2000a; Lyons & Chaterjee, 2002). Traffic flows were considerably affected by the scarcity of fuel, with congestion extremely low in the first week of closure, to the extent that any changes could not be attributed to the bridge closure; neither had our design anticipated how to survey the impacts of the fuel shortages. We thus re-arranged our surveys to monitor more closely the planned re-opening of the bridge. Unfortunately these surveys were hampered by a second unanticipated event, namely the wettest autumn in the UK for 270 years and the highest level of flooding in York since records began (BBC, 2000b). The flooding closed much of the centre of York to road traffic, including our study area, as the roads were impassable, and therefore we abandoned the planned 'after' surveys. As a result of these events, the useable data we had (not affected by the fuel protests or flooding) consisted of five 'before' days and one 'during' day.

In the Fishergate study, fortunately no extreme events occurred, allowing six 'before' and seven 'during' days to be surveyed, together with one additional day in the 'during' period when the works were temporarily removed. However, the works over-ran into the long summer school holidays, when it is well-known that there is a substantial seasonal effect of much lower flows and congestion levels. We did not believe it possible to meaningfully isolate the impact of the link fully re-opening while controlling for such an effect, and so our plans for 'after re-opening' surveys were abandoned.

3. ESTIMATION OF VEHICLE MOVEMENTS & TRAVEL TIMES

The data resulting from the surveys described in section 2 is in the form of (for each day and each study) a set of time-stamped, partial licence plates, observed at a number of locations across the network. Since the data include only partial plates, they cannot simply be matched across observation points to yield reliable estimates of vehicle movements, since there is ambiguity in whether the same partial plate observed at different locations was truly caused by the same vehicle. Indeed, since the observed system is 'open'—in the sense that not all points of entry, exit, generation and attraction are monitored—the question is not just which of several potential matches to accept, but also whether there is any match at all. That is to say, an apparent match between data at two observation points could be caused by two separate vehicles that passed no other observation point. The first stage of analysis therefore applied a series of specially-designed statistical techniques to reconstruct the vehicle movements and point-to-point travel time distributions from the observed data, allowing for all such ambiguities in the data. Although the detailed derivations of each method are not given here, since they may be found in the references provided, it is necessary to understand some of the characteristics of each method in order to interpret the results subsequently provided. Furthermore, since some of the basic techniques required modification relative to the published descriptions, then in order to explain these adaptations it is necessary to understand some of the theoretical basis.

3.1 Graphical method for estimating point-to-point travel time distributions

The preliminary technique applied to each data set was the graphical method described in Watling & Maher (1988). This method is derived for analysing partial registration plate data for unidirectional movement between a pair of observation stations (referred to as an 'origin' and a 'destination'). Thus in the data study here, it must be independently applied to given pairs of observation stations, without regard for the interdependencies between observation station pairs. On the other hand, it makes no assumption that the system is 'closed'; there may be vehicles that pass the origin that do not pass the destination, and vice versa.

While limited in considering only two-point surveys, the attraction of the graphical technique is that it is a non-parametric method, with no assumptions made about the arrival time distributions at the observation points (they may be non-uniform in particular), and no assumptions made about the journey time probability density. It is therefore very suitable as a first means of investigative analysis for such data. The method begins by forming all pairs of possible matches in the data, of which some will be genuine matches (the pair of observations were due to a single vehicle) and the remainder spurious matches. Thus, for example, if there are three origin observations and two destination observations of a particular partial registration number, then six possible matches may be formed, of which clearly no more than two can be genuine (and possibly only one or zero are genuine). A scatter plot may then be drawn for each possible match of the observation time at the origin versus that at the destination. The characteristic pattern of such a plot is as that shown in Figure 4(a), with a dense 'line' of points (which will primarily be the genuine matches)

superimposed upon a scatter of points over the whole region (which will primarily be the spurious matches). If we were to assume uniform arrival rates at the observation stations, then the spurious matches would be uniformly distributed over this plot; however, we shall avoid making such a restrictive assumption.

The method begins by making a coarse estimate of the total number of genuine matches across the whole of this plot. As part of this analysis we then assume knowledge of, for any randomly selected vehicle, the probabilities:

$$\theta_k = \Pr(\text{vehicle is of the } k^{\text{th}} \text{ type of partial registration plate}) \quad (k = 1, 2, ..., m)$$

where

$$\sum_{k=1}^{m} \theta_k = 1 \ .$$

The particular methods for estimating these probabilities in the case-study examples will be described in section 3.3. It is noted that in the original reference of this work, the simplifying assumption of equally-probable types was made throughout, i.e. $\theta_k = \frac{1}{m} \ (k = 1, 2, ..., m)$, but the analysis is readily generalised to the more general case considered here (see below).

Now, even if the streams of vehicles passing the origin and destination were entirely independent, we would still by chance get a number of possible matches, all of which would clearly be spurious. Given a total number of vehicle observations (disregarding the registration information) at the origin and destination respectively of n_1 and n_2, then there are $n_1 n_2$ pairs of observations that could match in this way 'by chance'. But, under the assumption of independent traffic streams at the observation points, the probability of any randomly selected pair of observations having the same partial registration plate type, purely by chance, is:

$$\Pr(\text{spurious match, if streams independent}) = \sum_{k=1}^{m} \theta_k^2 \ .$$

Multiplying this probability by $n_1 n_2$ gives the expected number of 'chance' matches (namely $n_1 n_2 \sum_{k=1}^{m} \theta_k^2$) under the hypothesis of independent traffic streams. In any real study (where we hope the independence assumption will fail, i.e. some vehicle pass both observation points), this figure may then be used as an estimate of the 'background level' of spurious matches; subtracting this from the number of possible matches therefore gives an estimate of the number of genuine matches. However, this is a relatively crude method since we have treated the data en masse; since any point on the scatter graph implies a 'journey time', this suggests that we should be able to deduce an improved method that essentially exploits the plausibility of these implied journey times in different parts of the scatter plot (clearly negative values for the implied journey times could easily be ruled out for example, but our approach can go beyond this).

Applying this same principle, therefore, to sub-sections of the scatter plot of possible matches, then for any given value of y (representing a journey time), we can focus on a triangular section of the plot bounded by the edges of the graph and the inequality $t - s \le y$, where the s and t respectively denote observation time at the origin and destination. Within such a triangular region, we can subtract the background level of spurious matches from the number of possible matches, yielding an estimate of the number of genuine matches with a journey time $\le y$. Dividing this figure by the estimated number of genuine matches over the whole plot (as derived above), we obtain an estimate of the proportion of genuine matches with a journey time $\le y$. In this way, we are building up an empirical estimate of the journey time cumulative distribution function for genuine matches,

$\hat{F}(y)$ $(-\infty < y < \infty)$. Clearly this can be used to deduce an estimate of the genuine journey time probability density, $\hat{f}(y)$ $(-\infty < y < \infty)$.

In simulation experiments, reported in Watling & Maher (1988), $\hat{F}$ was seen to be an 'improper' estimator, in that it is not guaranteed to be non-decreasing, and is not confined to the range [0,1]. On the other hand, there was seen to be a 'trust region', safely between the 20th and 80th percentile, within which the method was seen to recover reliably the overall shape of the journey time densities specified to the simulation experiments.

3.2 Maximum likelihood estimation of no. of genuine matches and journey time parameters

While the graphical method is attractive as a first line of investigation, due to its limited assumptions, there have been found to be methods with much greater potential precision for point estimation, in particular for estimating quantities such as the number of genuine matches and moments of the journey time density (Watling, 1990), specifically the maximum likelihood method (Watling, 1994). This also has the advantage over the graphical method of being able to simultaneously analyse data from multiple entry and exit sites to the study area (we shall henceforth refer to these as 'origins' and 'destinations'), rather than having to match origin-destination pairs independently or in some priority order. Clearly, by its name, the maximum likelihood method is a parametric technique, and its success depends on the validity of the distributional assumptions made. In extensive simulation trials, the method was seen to be robust to all assumptions (such as errors in the θ_k probabilities) except for that of the assumed distributional form of the journey time density function of genuine matches (Watling, 1990).

The assumptions made are simply that: at each origin, traffic splits as a multinomial distribution between the available destinations (including a 'sink' destination, representing vehicles observed at an origin which do not travel to any of the observed destinations); the journey times between each origin-destination pair are Normally distributed; for each destination, before arriving at the destination, the combination of traffic from all origins is joined by traffic from a nuisance 'source' origin (representing vehicles observed at a destination that were not observed at any origin), which arrives according to a time-homogeneous Poisson process. The estimation task is then to estimate the parameters of this model, namely the multinomial split probabilities for each origin, the mean and variance of the Normal journey time distribution for each origin-destination movement, and the arrival rate of the nuisance source Poisson process for each destination.

Allowing for ambiguities in matching the data, a likelihood function may then be deduced, though this cannot be maximised analytically. A direct maximum likelihood technique thus involves numerically maximising a likelihood function with an enormous number of terms, relating to all the possible complete matchings of the observed data. Even with efficient gradient-based optimization algorithms, this was seen to limit the practicality of the method for large-scale data sets. Therefore, an alternative method was proposed, based on a general purpose statistical algorithm for maximum likelihood estimation in the presence of 'incomplete' data. This method, known as the EM algorithm, operates by alternately (i) (E-step) computing the expected complete data log-likelihood based on the current estimates of the parameters from step (ii); (ii) (M-step) forming estimates of the parameters by maximising the expected complete data log-likelihood from step (i). Each of these steps can be performed analytically, and are performed alternately until convergence.

In simulation experiments (Watling, 1994), the algorithm was seen to converge reliably, recovering parameter estimates with high precision (notably comparable in magnitude to the precision obtained

by direct tracking of vehicle movements, with no matching ambiguity). It was found, however, that bias could arise in the parameter estimates in two circumstances, as described below.

Firstly, the presence of end effects may cause a bias most seriously in the multinomial split estimates, but also in the journey time parameters. Such effects arise due to the finite observation time intervals, meaning that a vehicle recorded within the observation period at an origin may travel to a destination but arrive outside the observation period there[5]. These are termed 'end effects' as they are most serious for vehicles observed towards the end of the origin observation periods. The impact is a systematic underestimation in the proportion of origin-vehicles that travel to each of the destinations. A secondary impact is a potential underestimation in the journey time mean and variance, since the distribution is 'right-censored' by the destination period: vehicles with longer journey times are more likely to be missed. While a theoretical fix to the method is possible to overcome this problem (see Appendix, Watling, 1994), this is at a relatively high computational cost, as the resulting M-step of the EM algorithm is no longer analytically solvable, and numerical optimization must be used. The problem was seen instead to be more readily overcome at the experimental design stage, by adjusting the relative sizes of the observation periods: see the discussion of §2 and analysis of §4.1.

Secondly, if the assumed journey time distribution is a poor fit to the shape of the actual data, then the journey time parameter estimates may be poor, and this can have a knock-on effect to the estimates of the multinomial split probabilities (and hence the estimated number of genuine matches). This issue was addressed in the present study by first using the graphical method as an investigative technique to deduce an appropriate distributional form. This also required a (minor) generalisation of the original maximum likelihood matching method, which was based on an assumption of Normal journey times, to allow parameters of alternative distributional forms such as the lognormal to be estimated.

3.3 Estimation of registration plate type probabilities

An important input to all of the registration plate analysis methods described above—as it affects the chance of a spurious match—is the estimate of the probabilities:

$$\theta_k = \Pr(\text{vehicle is of the } k^{th} \text{ type of partial registration plate}) \quad (k = 1,2,...,m).$$

A simple estimate can be made based on the assumption that all types are equally likely: we shall refer to this as the equi-probable estimate. In order to understand this issue better, we need to understand the system of assigning registration plates in the UK. In fact there have been several changes to the system of registering plates, meaning that at any one time one will observe vehicles corresponding to several systems. In particular the surveys were all conducted in a period where a now-defunct system of registration plate assignation was in operation. This meant that in the surveys to be analysed, the vast majority of partial registration plates were of the form 'Xijk', where the X can be one of 21 letters of the alphabet denoting the year of registration of the vehicle and the 'ijk' is a numerical code of up to three digits. Assuming all combinations of year letter and numerical code are theoretically possible to arise therefore gives 21000 combinations, i.e. possible 'types' in the terminology of the present paper. In fact, while this system described the vast majority of possible 'types', there are also a small number of partial registration plates that do not, such as some very old vehicles following an older registration system, non-UK registration plates,

---

[5] End effects also occur when vehicles observed toward the beginning of a 'destination' survey period passed an 'origin' point before the start of the period there. Simulation experiments indicated the main impact of such cases to be a bias in the estimates of the mean arrival rates of the nuisance Poisson sources (which are of no direct interest), with little impact on the main parameters of interest. Hence they are not considered further in this paper.

and some personalised number plates. It is difficult to estimate the true prevalence of such possibilities, so as an approximation we assumed that these in total amounted too a 1000 more possible types, giving 22000 possible types in total. Thus, the equi-probable estimate used was:

$$\hat{\theta}_k = \frac{1}{22000} \quad (k = 1,2,...,22000) \ .$$

As an alternative to the equi-probable estimate, an empirical estimate was made from the available data for the Fishergate study. While in principle this would be possible for all year-letter and number combinations, a problem is that if we pool all data from all sites then there will be a natural bias (double-counting) toward vehicles passing more than one site; this is not a trivial issue since the whole purpose of these estimates is to distinguish the chance of a genuine or spurious match. If, on the other hand, we use only data from entry sites, then we will have many types observed only at exit sites for which we have no frequency information. The approach adopted, then, was to make an estimate of the year-letter frequencies, pooling all entry sites, while assuming the three-digit numerical components to be equally likely. The frequencies of year letter can be affected by several things, on the whole one would expect a decay going back in time as vehicles are scrapped, but on the other hand (local and global) economic factors will affect the rate at which new cars are purchased. The empirical distribution estimated from the data is given in Figure 3.

4. ANALYSIS OF STUDY DATA (OVER DAYS)

The methods described in section 3 were applied to each study site and to each day of data in turn. In the present section, we shall explore the combination of all such analyses over all study days.

4.1 Exploratory analysis of journey time distributions

As a first step in the exploration of the data, the graphical method (described in §3.1) was applied to pairs of observation points for individual days in the survey period. As this method does not account for the interdependence in matching between anything more than two observation stations, the analysis is intended more for exploration, and to inform the subsequent parametric analysis. For illustration, we shall focus on the Fishergate study, and particularly on the movement most affected by the intervention, namely that from E to A. Figures 4(a)–4(d) give the results for this movement on the last day before the intervention (2nd July). Figure 4(a) shows the raw scatter plot of possible matches. Figure 4(b) presents the same information as in the scatter plot, but allows us to examine how the apparent journey times vary by observation time at E (i.e. the presence of any 'dynamic effects'). Figure 4(c) presents the estimated cumulative distribution function $\hat{F}$ corresponding to genuine matches. Finally, Figure 4(d) translates $\hat{F}$ into the form of an estimated journey time density function $\hat{f}$.

Consistently with the simulation experiments reported in Watling & Maher (1988), Figure 4(a) reflects (here for the real data) the characteristic clustering around the line $t - s = \mu$, where μ is the mean journey time of genuine matches, and (s,t) is a pair of observation times at E and A respectively. Since the observation periods at both sites finish at the same time, some end effects must occur. From the upper right corner of the plot, it appears that this relates to vehicles passing E after time 550 (i.e. in the last ten minutes of the survey), since this is where the trend line is censored, but it would appear to affect relatively few vehicles. From Figure 4(b), since the trend line appears relatively flat, there appears to be no appreciable temporal effects within the survey period (such as a within-day change in journey time distribution). Figure 4(c) presents the estimated journey time cumulative distribution function, which appears to have a legitimate shape over the 'trust region' of $0.2 \leq \hat{F} \leq 0.8$ noted in §3.1. Figure 4(d) illustrates the derived journey time

density function over this trust region. Qualitatively similarly results were found for other movements over both days before the intervention and during the suspension day (figures not shown here due to space limitations): that is to say, a similar general shape of the density function, similar structure to the scatter plot, and the lack of any appreciable temporal effects.

A more interesting comparison is found by examining plots for this same movement (E to A) for the first two days during the intervention, namely 3rd and 4th July, respectively shown in Figures 5(a)–5(d) and Figures 6(a)–6(d). Obviously, caution needs to be exercised in drawing more than indicative findings from this comparison, as it only involves a small number of days, but it does suggest important qualitative features to guide the further analysis. In particular, there are a number of striking contrasts between the plots for the intervention days and those without (not all shown here). Considering Figure 5(a), the dense trend line is more broken than in the 'before' case, the location of this line has shifted (indicating a location shift of the journey time distribution), and there is a much greater problem with end effects impacting on the last thirty minutes or more of the survey period at E. From Figure 5(b), there appears to be a build-up in journey times as the survey period progresses, as demonstrated by the slope of the dense trend line relative to the horizontal. Of particular relevance for the later parametric analysis, it is apparent from Figures 5(c)–5(d) that there has been a considerable change (relative to the 'before' situation) in the location, dispersion and shape of the journey time distribution, with a strong positive skew and much greater dispersion now evident. In Figures 6(a)–6(d), the evidence is similar, though with less apparent skewness in the distribution.

For reasons of space, we do not present here the equivalent graphs for other movements, but note that they did not demonstrate the same striking changes in the shape of the distribution as for the E to A movement. In particular, detailed attention was paid to two movements which, by their location, might also have been expected to be significantly affected, namely the E to K and G to A movements, with the resulting approximately symmetric journey time distributions for the two 'during' days resembling those for the 'before' and 'suspension' days.

Overall, the main suggestive evidence of this exploratory data analysis to take forward to the parametric analysis, were that a symmetric distribution of journey times appears to be a justifiable approximation for the vast majority of movements, but that for the journey E to A during periods of disruption there is evidence to suggest that a skewed distribution may on some days be more appropriate. Also, there is evidence of 'end effects', again especially for days during the intervention, which may need to be considered for their potential biasing effect on parameter estimates.

4.2 Parameter estimation for individual days & Between-day comparison tests

The exploratory analysis in §4.1 was able to make independent comparisons between pairs of sites, but did not consider the interdependencies between sites in the matching process. In any case, more (statistically) efficient parametric estimation methods exist, as noted earlier. Thus, for the purpose of allowing for the interdependencies between sites and of obtaining more precise estimates of summary measures, the maximum likelihood method described in §3.2 was applied. This was done for the Fishergate data across all sites and for each survey day in turn, and to selected days of the Lendal Bridge data set. The empirically-investigated registration plate type probabilities (see §3.3) were used throughout this analysis, though in tests it was found that the alternative assumption of equi-probable types did not have a major impact on the results; this is consistent with earlier simulation-based findings about the relative insensitivity of the method to these estimates. Starting values for the EM algorithm may be found by reference to the results of the graphical method study in §4.1: as noted in §3.1, the denominator of the estimator of the journey time cumulative

distribution function is effectively a coarse estimator of the number of genuine matches between a pair of points, and this may be divided by the total number of vehicles observed at an origin to obtain an initial estimate of the relevant multinomial split probability. This information may also be used to obtain an estimate of the number of nuisance source vehicles at each destination, and divided by the length of the observation period to get an estimate of the relevant Poisson rate. Finally, initial journey time estimates may clearly be made from the summary statistics of the density functions estimated by the graphical method. It is noted, however, that our tests here (not shown) confirmed earlier simulation results, that the EM algorithm final results were relatively insensitive to the choice of starting condition, so the precise manner of obtaining such initial estimates is not so important. Our discussion below will primarily focus on the Fishergate study, due to the better data quality in that case (as explained earlier), but we shall conclude by examining the Lendal Bridge study data to examine transferability of the qualitative findings to other sites.

According to the exploratory data analysis of §4.1, Normal journey times were assumed in the maximum likelihood method in all cases, except for movement E to A during the capacity reduction, when a three-parameter log-normal was adopted. Table 1 contains selected results of this analysis for the Fishergate data in terms of the estimated number of genuine matches, and journey time mean and standard deviation, across all survey days. The matching is given for the most significant movements (in terms of their potential to be affected by the capacity change), namely E to A, E to B, E to K, E to F, G to A and G to B, where the survey locations are depicted on Figure 1. The columns of Table 1a are, working from the left, the date, the number of genuine matches for the four movements from E for that day, the total observed flow at E for that day (irrespective of whether it matched with any of the destination sites), the percentage of the flow at E which was matched with one of the destinations, and then similar information for the movements from point G.

For the purposes of an initial analysis it was decided to test between two groups of days: the five 'before' days, preceding July 3rd, and the six 'during' days where the capacity reduction was in operation, excluding the final two survey days (the temporary suspension day of July 13th and the day following it). The mean matched flow over the six during days (which can be calculated as 247.7) is substantially less than the mean over the five before days (297.6). A two-sample, unequal variance t-test for a reduction in the mean yields a one-sided p-value of p = 0.000872, and so this apparent reduction is highly significant.

This reduction in the EA flow could be due to a number of reasons; since the location of the capacity reduction means that this will be a major affected route, we investigate the potential reasons in some detail below:

1. Firstly, it might be that due to some external factors, the 'during' days were particularly 'low flow days' generally. However, performing the same analysis as above on the mean total flow at E, then there is in fact an observed increase during the capacity reduction (from 1427.4 to 1495.8), though this is hardly significant (2-sided test, p = 0.092254)[6]. Thus, there is certainly no evidence that the total E flow is the cause of the reduction in the EA flow. It is noted that this also apparently rules out 'strategic' re-routing, in the sense of there being no evidence of a decreased propensity to choose routes that pass through E as a result of the capacity reduction.

2. Secondly, it might be a 'capacity' effect, in the sense that a reduced throughput over a fixed time interval would be reflected in a decreased matched flow. In particular, it should be recalled that a substantial impact was observed when applying the graphical method in section 4.1 to the 'during' data for EA, resulting in significant problems with 'end effects'. To explore the impact of the end effects, the EM algorithm was re-applied to a reduced form of the data. This was

---

[6] The subsequent end-effect corrections, described later in the section, only further highlight the non-significance of the change in total E flow: with the increase then from 1099.4 to 1164.2, at a two-sided value of p = 0.154348.

achieved by discarding origin data at point E that departed in the period after 8:55, that is to say discarding the last 20 minutes of the origin survey, but matching against the full survey periods at the other sites. The reason for choosing the time 8:55 is illustrated in Tables 2 and 3, whereby the truncation period at E has been altered over a range, and the impact on parameter estimates examined. For the 26th July, reductions below 9:10 begin to show an impact on the estimated parameters, with the flow decreasing, as matches are "lost". On the 3rd July, however, a large reduction in the survey period is possible, down to less than 8:55, before the parameters begin to change.

Table 4 provides the counterpart to Table 1 for this truncated data set (i.e. with the 'origin' site truncated at 8:55). Clearly, as would be expected, the absolute number of matches are smaller than in the untruncated case. However, it can be verified that the truncation is successful in appreciably increasing the percentage of origin vehicles that are matched with some destination vehicle.

Repeating the test on the mean EA flow before and during the capacity reduction, the reduction from 250.0 to 237.3 gives a one-sided p-value of $p = 0.033699$. That is to say, the significance has been greatly reduced by the truncation, indicating that a major factor in the difference of the original (untruncated) EA matched flow was indeed a capacity effect. However, the difference is still highly significant, and so this is not the only explanatory factor.

3. Thirdly, there may have been a diversion away from EA, either because of advanced publicity of the road maintenance, en route diversion, or a 'learning effect'. In particular we examine whether this could be the explanation for the remaining significance in the EA mean flow reduction, after the truncation described in point 2. above has been applied. This is achieved by examining, for each day, the proportion of the total vehicles observed at E that are EA-matched journeys, and then testing the change in the mean proportion between the 'before' and 'during' days. This reveals a decrease in the fraction of drivers passing E that follow path EA, from 0.228 to 0.204, which (while it might appear small) is highly significant at a one-sided $p = 0.000477$. The reason for being able to detect significance with apparently such a small mean change is the remarkably small between-day variance in flow proportions for each of the two scenarios individually: the actual proportions being 0.233, 0.242, 0.228, 0.221, 0.216 before the capacity reduction; and 0.207, 0.200, 0.196, 0.213, 0.208, 0.200 during the capacity reduction. (In other words, the between-day variations in the EA flow are, taking either the 'before' or 'during' scenarios in isolation, most strongly explained by between-day variations in E, not by between-day variations in the fraction of E drivers selecting EA). It is noted, however, that one cannot discern any 'learning' effect, in the sense that the 'during' days are stable from the first day (perhaps an indicator of good advance warning of the works). In conclusion, it can be said that the capacity reduction certainly appears to have had a significant effect on the fraction of drivers passing E that select path EA.

Following a similar analysis for another of the major movements, EK, and making the truncation to remove end effects, a significant increase in the mean EK flow (from 233.0 to 285.8, one-sided test for an increase: $p = 0.002056$) can be seen in comparing the 'before' and 'during' data. Furthermore, when comparing the proportions of E flow following EK, there is again remarkably little between-day variation within either the 'before' or 'during' periods, but a highly significant increase in the mean proportion between the periods (from 0.212 to 0.246 at $p = 0.00097$). The analysis of flow proportions on EA and EK therefore suggests the data are consistent with a diversion away from EA to the alternative route EK, though we cannot attribute this with certainty.

As an illustration of the intervention's impact on other movements (detailed figures not given due to space limitations), let us also consider movements emanating from site G. In particular, analysing

the major movement GA, no significant mean change was detected at 5% significance in either the GA absolute flow, the G total flow, or the GA flow proportion. Examining the proportional flow GA of G over days, one obtains 0.406, 0.426, 0.419, 0.402, 0.413 before the intervention, and 0.443, 0.407, 0.376, 0.384, 0.398 for the days during the intervention. Again the proportions are quite stable within the before or during periods, except that the first day of the capacity reduction appears to stand out as slightly unusual: it has a substantially higher flow proportion than other days (and the lowest total flow at G). It might be postulated that this may be consistent with a learning effect, and if one could justify removing this day as unusual, then one does obtain a significant reduction in the mean GA proportions between the before and during cases (one sided test for a reduction, p = 0.022533).

Turning attention now to the journey time statistics, the first question is whether to base the analysis on Table 1 or Table 4. End effects have the potential to cause a bias in journey time statistics too, since they will tend to disproportionately affect longer journey times, leading to a potential underestimation in the journey time mean and standard deviation. However, examining the journey time results in Tables 1 and 4, there is no particularly clear evidence of this effect, and both sets of tables appear to show similar trends. Therefore, for the purposes of consistency with the graphical matching analysis (which was based on the full, untruncated data), and in order to maximise the sample size, we shall focus on Table 1, though the findings below would be qualitatively similar had we adopted the figures from Table 4.

We consider as an example again movement EA, and focus specifically on the first five survey days, before the capacity reduction. From Table 1, there is evidently quite substantial between-day variation in the mean journey time. Thus, even if we were to detect a change in the mean journey time between (say) the day before and day after the capacity reduction, this could not be confidently attributed to be an impact of the capacity reduction. However, treating each of the day-specific mean journey times as observations, we can compare the between-day mean of these means in the scenarios before and during the capacity reduction. In this comparison, the between-day mean shows an observed increase from 7.57 minutes to 10.28 minutes, which—even in the light of the substantial between-day variation noted within each scenario—is significant at p = 0.0118.

Having made the analyses and observations above based on the Fishergate study, we look to the Lendal Bridge data for the presence of corroborating or conflicting evidence with the qualitative findings above. Although the data are more limited (due to the problems described in section 2), they do appear to give some interesting corroborative insights. Looking to Figure 1, we particularly focus on the flows over the three bridges—Lendal (site H), Ouse (site J) and Skeldergate (site C)—which may be viewed as alternative river-crossing locations for flows from Blossom Street (site A, in the west) for traffic destined anywhere east of the river.

The results of the matching analysis from the EM algorithm are given in Table 5, where we specifically focus on the estimate of the number of genuine matches. As noted earlier (see §2), the first five days represent the 'before' data, which are 'clean' in that they were unaffected by either the fuel protests or floods. The final day (27/9) represents a clean day of 'during' data, when Lendal Bridge was closed. It would be natural to seek to apply a similar kind of analysis to that applied for the Fishergate study above, in particular computing p-values for a null hypothesis of no-change. However, a difficulty here is that due to the problems noted earlier in the Lendal Bridge study, we have only a single day of data for the during-intervention case, and so we have no opportunity to account for ambient day-to-day variability during the intervention. Instead, then, we consider the figures in a slightly different way, by first examining the data before the intervention (i.e. excluding the 27/9 day), and then making a qualitative comparison with the 'during' data.

In particular, we may form 95% confidence intervals for the true mean flow from Blossom across each of the bridges, before the intervention, which yields confidence intervals (to the nearest integer) of [118, 156] for Lendal, [65,118] for Ouse and [89,205] for Skeldergate Bridges. These are notably rather wide confidence intervals, especially for Skeldergate, reflecting the considerable between-day variability observed, and in such a case it is perhaps not surprising that our single day of 'during' data does not provide evidence of a significance change in the flow over this bridge (as the single day of data, even neglecting the variability in the 'during case', is already within the 95% confidence interval for the 'before' data). For the Ouse bridge the decision is more marginal, with the 'during' observation on the boundary of the 95% confidence interval for the 'before' data, but had we allowed for the variability in the 'during' data then certainly here too we would have no evidence of a statistically significant change in flow. However, it is notable that one reason for this difficulty is the considerable day-to-day variability in the total flow from Blossom Street, as seen in Table 5.

This suggests adopting a similar strategy to that used in the Fishergate analysis, namely analysing the flow proportions over the bridges (i.e. as a proportion of the flow at site A). Forming 95% confidence intervals for the true mean proportions of flow at A choosing the different bridges, before the intervention, gives intervals of respectively [0.172, 0.200] for Lendal, [0.088, 0.160] for Ouse, and [0.153, 0.234] for Skeldergate. On the intervention day the proportion observed at Ouse was 0.213 and at Skeldergate 0.257, both outside the 95% confidence interval for the 'before' data. While we cannot conclude that this gives statistical significance without evidence of the day-to-day variability during the closure, it provides a qualitative indication of a potentially significant impact. This is not so surprising, given that we have closed one of three bridges and seen a change to the traffic choosing the two obvious alternative bridges; the main issue concerns the corroboration of the analysis of the earlier Fishergate data, where we were able to see considerably more chance of significant information arising from analysing flow proportions as opposed to absolute flows.

5. COMPARISON WITH NETWORK EQUILIBRIUM MODEL RESULTS

The final step of the study involved comparing the empirical data with the predictions of a traffic network equilibrium model. There are many reasons that make such a comparison less than straightforward, and particular issues that need to be borne in mind. Firstly, the question is what kind of network equilibrium in particular do we want to use, such as dynamic or static, deterministic (Wardrop) or stochastic user equilibrium, route choice based on mean travel times or on unreliability considerations, etc. In fact we decided to approach this problem from a practice-based perspective, and particularly what methods were commonly used in practice. This in fact opens up further model classes, which one does not often see considered in the academic literature, namely the models that have been developed for particular software packages that often lie somewhere between the distinctions seen in the academic literature.

In particular, we chose to use the so-called 'simulation-assignment' capabilities incorporated in the SATURN suite (Van Vliet, 1982), based on a representation of an asymmetric, static Wardrop equilibrium problem with junction interactions. This technique pays particular attention to the detail of junction delays in urban networks, using an 'exploded' network definition in which each turn is represented by a link with its own flow-delay relationship. These flow-delay relationships are, however, not pre-specified, but are internally estimated during the assignment procedure, by fitting curves to points generated by a 'simulation' model. The word 'simulation' is perhaps misleading: in fact traffic flow propagation through and between junctions is handled through so-called cyclic flow profiles, time-sliced flows entering and leaving the links (turns and junction-to-junction links), disaggregated typically into a resolution of a few seconds, but with an assumption of cyclical behaviour during the whole modelled period that is assumed imposed by the dominant cycle time of

the traffic signal intersections. This flow propagation approach allows the method to overcome to some extent, albeit in a heuristic way, the problem of static traffic assignment models failing to impose capacity constraints on flows, with unserved demand held over on a link to be served in a subsequent modelled time period. The algorithm as a whole has the semblance of a diagonalisation algorithm: at one iteration of the diagonalisation, conditional on opposing/interacting flows as estimated from the previous iteration, separable flow-delay relationships are estimated for each link (turn) and a Frank-Wolfe algorithm is applied to find a Wardrop equilibrium route pattern conditional on those relationships. This route pattern is then fed into the subsequent iteration of the diagonalisation, where a new 'simulation' is performed, new flow-delay relationships estimated, and a full Frank-Wolfe algorithm again applied. This heuristic method has no guarantee of convergence, though extensive use of it in practice indicates that it usually stabilises, except in highly congested networks.

We have given the details above for completeness, as they are not so readily found in the public domain, and because we felt that the method – while limited by the fact that it uses a static rather than dynamic user equilibrium – overcomes at least in a coarse way many of the concerns the reader might have with adopting a static approach. In fact our choice of method was also guided by the fact that a well-maintained, SATURN-based representation already existed for the City of York, and the fact that the project team had some considerable familiarity and experience with the software. Other software packages or other modelling methods could have been chosen, and in fact we shall not delve into the details of how well each of the elements of SATURN performs, as we wish to gain more general, qualitative findings about the model fit that are not specific to a particular software package. We believe that our qualitative findings below would hold good had other network equilibrium approaches been adopted, though we have not tested this claim.

Aside from the issue of which traffic assignment modelling representation/paradigm to choose, any comparison between a network equilibrium model and empirically-observed networks must also be made subject to several other considerations. Perhaps most significant among these is the issue of data quality: how good and up-to-date are the origin-destination matrix and the input network data? A quite different issue is the question of which numerical algorithm was used to implement the chosen paradigm: in our study we have used the Frank-Wolfe algorithm, for which we need to accept a likely larger convergence error than with several newer algorithms for solving Wardrop user equilibrium (e.g. Dial, 2006; Bar-Gera, 2010; Nie, 2010). While this in turn means a risk that part of the difference in any modelled comparison is attributable to such convergence error, we feel that the systematic nature of the results from the tests we report below provides evidence that, for the size of changes we have considered, the effects of changing the input parameters outweighs any convergence error (which is unlikely to lead to such systematic effects). In any case, we feel that such concerns about convergence error need to be weighed against those of the underlying model error, such as is caused by using of a single OD matrix and a single point equilibrium to approximate what is in fact a system with considerable real day-to-day variability.

The approach for comparing observed and modelled routes centred around the subset of observed results obtained from the Fishergate study that were reported in Table 1. The reason for choosing these results was that they contained the greatest, statistically significant impacts of the road capacity change, and so gave maximum potential for making a systematic comparison in the presence of the many kinds of error that exist (ambient daily variability, sampling error, model mis-specification error, convergence error, etc.). The comparison was used to test the hypothesis that the steady-state equilibrium flows and travel times from the assignment model could be used as approximations to the observed flows and mean travel times average over days, for each of the cases with and without the intervention (road capacity reduction). In fact in this comparison, the last two days of observed data in Table 1 have been neglected, when the restrictions were temporarily lifted for one day and then re-imposed, as we felt that they may contain atypical, transient effects.

We performed the comparison of observed and modelled information in two distinct stages:

1. A comparison of mean flows and travel times between pairs of sites in the base situation, prior to the intervention, and with Fishergate modelled at full capacity, to ensure that the model provided a sufficiently faithful representation of observed route choice behaviour to use as a test-bed.

2. A comparison of mean flows and travel times between pairs of sites for the "during intervention" days (those shaded in Table 1), with Fishergate now modelled at reduced capacity, to investigate the performance of the model in predicting changes in route choice behaviour.

It is worth emphasising the difference of this approach from 'model validation' as is commonly applied in practice. The practical approach usually adopted is (i) to only validate the model in the base situation, i.e. there is no step 2, in the sense that the impacts of past interventions on a given network are not used to inform the modelling of future interventions; and (ii) to apply step 1 only at the level of aggregate link flows and average link travel times, not at the finer level of resolution used here of point-to-point flows and point-to-point travel times. Thus the validation tests applied here adopt a much higher standard, in asking the model to reproduce changes to some systematic input, and to reproduce those changes at a point-to-point rather than single-site level.

From the model, therefore, we need to extract information not just at the link level, but at a finer level of resolution. A potential drawback of this, especially for the Wardrop user equilibrium model, is that we may then be asking for information that is not uniquely defined by the model itself. It is well known that under the assumptions commonly adopted, user equilibrium link flows are unique but the route flows are typically non-unique; our level of interrogation will be between the level of link and route flow, and therefore also gives rise to potentially a non-unique model solution at this level. To be strict, then, what we compare from the model is the final solution obtained by the application of a Frank-Wolfe algorithm starting from initial free flow travel costs. Recently, Bar-Gera (2010) has investigated the prevalence of such non-unique outputs, and particularly the tendency for some equilibrium solution algorithms to stop at extreme solutions, favouring particular origin-destination movements. In the case of the Frank-Wolfe algorithm, however, Bar-Gera found much less evidence for such extreme behaviour, and we believe that this further supports our selection of the Frank-Wolfe algorithm and our decision to use model outputs at a higher level of resolution than link flows.

Based on our initial comparisons for step 1. above, we found that the model produced mixed results across the network in terms of how well it reproduced the point-to-point flows and travel times in the base situation (we will clarify what we mean by this below). While this could be due to poor/unrepresentative origin-destination matrix or network data, we did not believe this to be the case here, as we knew that (from the model-side) some considerable care had been taken by the local council to maintain and update the model inputs over time, and also because (from the empirical-side) our survey had been carefully designed so as far as possible it did not reflect atypical demand or network conditions (see §2). While several other potential sources of error surely made some contribution to this effect (e.g. convergence error, non-uniqueness of model outputs, sampling error, model mis-specification error), our investigations revealed a systematic feature that we believed could be controlled for. In particular, there was a general tendency observed that, relative to the empirical data, the model under-predicted radial movements in favour of orbital movements. (It is noted in passing that we did not arrive at this observation 'by chance', it was a feature that members of the team had previously observed in other transport networks.) It transpires that this issue has such a substantial effect on the overall comparison that we shall now devote some considerable attention to it, and its causes in the model.

In particular, we shall focus on what turns out to be a key contributory factor to this feature within the model, namely the assumption in the network equilibrium model (one commonly adopted, in UK practice at least) that drivers based their route choice decisions on generalised costs, these generalised costs having two components: travel time and distance. Thus, the generalised cost c_r (in pence) of any route r is given by:

$$c_r = \alpha t_r + \beta d_r$$

where t_r (minutes) and d_r (km) are respectively the travel time and length (distance) of route r, and where α (pence per minute) and β (pence per km) are respectively the values placed on a unit of travel time and a unit of distance. Since the results of the model are invariant to a scaling of costs across all routes, we may divide the generalised costs by β (say) to make it clear that in fact there are not two degrees of freedom, but only one in choosing the ratio α/β (km/minute) .

Evidence for the value of this ratio may be drawn from several different sources and we consider two particular, common approaches here. The first approach is to interpret the valuation on a km of distance as a proxy for the valuation of vehicle operating cost per km, which means that economic-based surveys may be directly used to derive values of time and operating cost, through stated or revealed preference approaches. The second approach is to treat the ratio as a calibration parameter for the network model, and to adjust it in order that observed network data is reproduced in the base situation. Each of these approaches has their merits and drawbacks, and it is useful to be clear on these points as they will be pertinent to our subsequent discussion.

On the one hand, economic surveys are methodologically consistent with traditional approaches for evaluating the benefits of investment, so tend to be favoured by economists and policy-makers. However, the surveys on which these values are estimated rarely include spatial route choice scenarios, so rely on the assumption that values derived from other (non-spatial) transport contexts are transferable. Furthermore, in practice many assignment modelling applications are created without access to locally-derived economic data to set coefficient values. In the absence of local data, value from national data are used, but these may be systematically biased estimates of local circumstances (eg due to differences between the local and national mix of socio-economic attributes) and of the context of the trips made (eg traffic assignment models typically focus on peak-hour commuting trips).

On the other hand, direct calibration ensures that coefficients are chosen to best replicate observed local conditions in terms of flows and journey times, but these may still be relatively crude indicators of actual route choices, and the performance of the model will depend critically upon the accuracy of the network representation, which in many cases may be being calibrated in parallel. Furthermore, this in turn raises questions when a subsequent economic appraisal is conducted: is it justifiable to use behavioural values in the route choice model that differ from the values used in the economic appraisal (since the latter may be set down in national guidelines)?

Let us now turn attention to our particular case-study. This network has been set up originally assuming a ratio for the generalised cost parameters of 2 km/minute, partially based on a local stated preference study. It is a little difficult to understand the units of this ratio, and what it is implying, and so to aid in the discussion Table 6 has been created. The purpose of this table is to show, for a given average speed of travel, what the implication is of assuming a particular value for the ratio α/β in terms of its impact on the relative contribution of travel time and distance to generalised cost. The values in the table are simply derived from noting that if generalised cost

$c = \alpha t + \beta d$, with t in minutes and d in km, then the time component (αt) makes up a proportion of the whole generalised cost ($\alpha t + \beta d$) of:

$$\frac{\alpha t}{\alpha t + \beta d} = \frac{\alpha/\beta}{\alpha/\beta + d/t} = \frac{\alpha/\beta}{\alpha/\beta + s/60}$$

where s is the speed in km/h (the 60 divisor just scales speed to more familiar units). Table 6 gives the resulting value of this proportion, converted to a percentage by multiplying by 100, for a number of given values of the ratio α/β and the speed s. For the York network as a whole, the average modelled network speed[7] for the full study area, including the fast A64 orbital bypass route, was around 40 km/h, and so from Table 6 we see that assuming $\alpha/\beta = 2$ can be interpreted as saying that on average, the weighted travel time component makes up around 75% of generalised cost: that is to say, we might interpret this as assuming that travel time is a rather dominant factor in the choice of route.

Moreover, for particular travel movements this contribution (of the weighted travel time to generalised cost) will vary: the figure of 75% represents a network-wide summary. The focus of our comparisons will be on urban radial routes close to the city centre, which will have speeds lower than the network-wide average. For these movements, we can see from Table 6 that the contribution will be even greater; for example, weighted travel time makes up 92% of generalised cost at a speed of 10 km/h, assuming a relative valuation of $\alpha/\beta = 2$.

In order to set this value in some context, several previous studies (e.g. May & Milne, 2000; Sumalee, 2005) have adopted values based on UK economic data (DfT, 1989) which gives coefficients of $\alpha = 7.63$ pence/minute and $\beta = 5.27$ pence/km, implying a ratio $\alpha/\beta \approx 1.4$. At that time, UK national data did not differentiate at all by region, by road type (urban vs inter-urban) or by time-of-day. In addition, since most traffic assignment studies operate with a single composite user class, distinctions available in the data for vehicle type and journey purpose are typically not utilised. So, the values adopted were national all-day all-location means based on assumptions of typical traffic composition. While regional aggregation may be expected to impact upon both values of time and operating costs (potentially in equal proportions), the road-type, time-of-day, trip-purpose and vehicle-type aggregations would be expected to impact primarily on the value of time. It would therefore not be unreasonable to suggest that an appropriate value for this ratio for the a peak-hour traffic assignment study might be somewhat lower than this national average of 1.4., since in our present context the traffic composition may reasonably be expected to be skewed towards private car-based commuter trips at the expense of business traffic, with the latter usually considered to have the highest values of time. Indeed, from anecdotal evidence of discussions with practitioners, it would not be unusual to see ratios of around $\alpha/\beta \approx 1$ assumed. National economic data in the UK has subsequently become more sophisticated, adding differentiation by road-type and time-of-day as models of urban peaks have become the norm. However, the practice of using a single composite user class remains, so aggregation of vehicle type and journey purpose is still

---

[7]A complicating factor is that obviously this modelled speed varies with the assumed value of the ratio, since this ratio affects the choice between routes of different lengths and travel times, but the effect is not so great that we need to concern ourselves with it for the purposes of the qualitative discussion of Table 6, which is only intended to be indicative.

common practice. Even taking account of the greater differentiation that exists, coefficients based on national UK data still tend to be consistent with ratios in the range of 1.2 to 1.4.

Referring again to Table 6, we can see that for ratios in excess of 1 (as applies to all the examples quoted above) and for the slower speeds typical in congested urban peak periods (10-30 km/h), travel time makes up 70% or more of total generalised cost. The concept of Wardrop User Equilibrium assignment implies that route choice decisions are extremely sensitive to small changes in total generalised cost. As the time-related proportion of total generalised cost increases, so it follows that route choice will become increasingly sensitive to small changes in travel time and delay.

So what effect does this have on the modelled results for our Fishergate study? Table 7 reports the comparison for the observed and modelled flows and travel times, for the base scenario without the capacity reduction. A range of values for the ratio α/β has been tested, greater than the range suggested by the economic studies above. As can be seen from the top two lines of Table 7, there was a great discrepancy between the modelled and observed results, under the initial assumption of $\alpha/\beta = 2$, with (a) the total flow modelled as choosing routes through E being only 80% of that observed and (b) the two major movements EA and EK carrying only 42% and 19% of the traffic expected. Even reducing the ratio to $\alpha/\beta = 1$, as noted from the evidence above, only achieves a small benefit in bringing the modelled and observed values closer. On the other hand, treating the ratio as a calibration parameter, then we can see from the table that assuming considerably smaller values potentially allows us to better match modelled and observed flows. However such values for the ratio are considerably outside the range suggested by economic surveys: so what evidence might we have to support such an approach?

For this purpose, we further interrogated the traffic assignment model, in order to identify the origin and destination points of all trips which choose routes along the surveyed corridor, allowing us to find out which, if any, non-surveyed routes may also have been chosen. The first stage of this more in-depth route choice analysis involved identifying the potential for increased usage of the Fulford Road corridor within the model. A simple indicator for this was derived via a select link analysis to identify those O-D movements choosing routes through point E (as shown in Figure 1) in a given scenario, allowing a comparison with the volume of trips for the same O-D movements assigned to alternative routes. This showed that under the initial assumption of $\alpha/\beta = 2$, although the volume of trips predicted to travel through point E was just 880, the total number of trips in the SATURN matrix for the O-D movements identified was 1339. Thus, of the journeys for which Fulford Road was identified as a feasible route within the assignment model, 34% of trips were choosing to go other ways. Given the topology of the York road network, any trips avoiding E would potentially need to take a considerable detour around the outer orbital route, so this degree of multiple routing was a little surprising.

The second stage of the route choice analysis involved exploring the assignment outcomes for a selection of O-D movements that might be expected to use routes through site E. The left hand column of Figure 7 illustrates the routes identified in the model under $\alpha/\beta = 2$, for four journeys with origins (represented by green stars) to the south of point E, but destinations (represented by red stars) that are geographically dispersed. Links highlighted in red show chosen routes, while green bandwidths show the proportional magnitude of flows. In all cases, it can be seen that the most direct routes along the Fulford Road corridor (i.e. through point E and, in particular, movements EA

and EK) are largely avoided in favour of more circuitous alternatives chosen on the basis of lower travel times. While some diversion onto faster, longer orbital routes may be intuitively sensible in a congested peak period, the extent of the avoidance of Fulford Road does seem rather surprising.

On the basis of this analysis, the evidence for problems with the route choice model being the primary cause of the discrepancy between observed and modelled data seemed strong. It was, therefore, decided to attempt recalibration of the model through the α/β ratio. From the initial route choice plots in Figure 2, we might expect that ratios placing greater emphasis on travel distance would encourage increased use of Fulford Road, so a range of α/β ratio values was tested from 1 to 0.1. In addition, a distance-only assignment ($\alpha/\beta = 0$) was added as a limiting case. As noted above, the results of these tests are shown in Table 7; and for a particular case ($\alpha/\beta = 0.2$) the right-hand column of Figure 7 illustrates the corresponding routing patterns.

From Table 7 it may be noted that the volume of traffic through site E, and for the major movements EA and EK, increases steadily as the ratio α/β is altered so as to favour distance relatively more highly to travel time. The modelled flow EB is always zero since the spatial aggregation of zoning system means there is nowhere for local for trips to terminate in the model, and because this movement is always likely to be costlier than EK for through traffic. The modelled flow EF increases from a negligible back-route to a significant rat-run as the ratio is decreased. The comparison of modelled and observed travel times is also informative for calibration: in the initial situation, under $\alpha/\beta = 2$, the modelled travel times are much lower than those observed. This can be explained by the fact that, since in such a case the Wardrop Equilibrium is based largely on travel time, this causes drivers to be better at avoiding delays by finding the most efficient routes in terms of travel time minimisation. Only when the α/β ratio is in the range of 0.3 to 0.2 are the modelled travel times comparable to those observed, and this is also the range where the modelled flows fit best. On the basis of these comparisons, we would suggest that $\alpha/\beta=0.2$ provides the best balance between satisfying the flows and the travel times. Returning to the issue discussed earlier in relation to Table 6, we note that with $\alpha/\beta=0.2$ then the contribution of weighted travel time to generalised cost has fallen to 23% at the network-wide average speed of 40 km/h (compared with 75% when $\alpha/\beta=2$ was assumed), and to 55% at a speed of 10 km/h (compared with 92%).

The right hand column of Figure 7 shows the routing patterns for this preferred case of $\alpha/\beta=0.2$, for the same selection of O-D pairs investigated under $\alpha/\beta=2$ in the left hand column. In three of the cases under $\alpha/\beta=0.2$, a route along Fulford Road which previously (under $\alpha/\beta=2$) carried only a small proportion of the total O-D flow has become the dominant route. In the one case where the $\alpha/\beta=2$ assignment produced no route that travelled the length of the Fulford Road corridor, the majority of the O-D flow in the preferred case still avoids it, but nevertheless there is now a minority route following the shortest path through the centre.

Thus overall there is good evidence that a major part of the discrepancy between modelled & observed site-to-site flows and travel times can be eliminated by 'calibration' of a single parameter controlling the ratio of values of travel time to distance. Following the identification of the preferred scenario (with $\alpha/\beta=0.2$), we believed it was then reasonable to use the traffic equilibrium model as a basis for predicting the effect of the intervention on Fishergate. As noted in §2, this intervention involved the closure of a lane on Fishergate, meaning that the turns to sites A and B in Figure 2 needed to share a single lane during the intervention, whereas they had a separate lane each prior to the intervention. Since the internal SATURN traffic model deals with the competition effects of lane-sharing on the capacities of the two turns, the intervention is represented by a change to the input Saturation Flow (SF) for each of the turns – defined as the theoretical maximum flow that could exit a turn in the modelled period if there were no competition from other turns. While an assumption of a 50% reduction in SF may at first seem reasonable, the frictional effects of

temporary narrowing may reduce the figure further than this, therefore a number of reductions in SF were tested. The results of this exercise are given in Table 8.

From Table 8 we can see that the modelled results during the intervention are quite sensitive to the SF assumption. The case SF=40% provides a reasonable ball-park estimate of the flows and travel times during the intervention, even though in absolute terms the fit of the flows in not as good as in the pre-intervention case (but this is not surprising, as we calibrated to that case). However, it should be borne in mind that there is an effect in the observations that the network equilibrium model will never be able to reflect, namely that the observed mean flow through site E actually increases during the intervention days. This is counter-intuitive to the logic of the equilibrium model, however calibrated, as there has been a decrease in capacity for routes passing through E, meaning that travel times must increase and so traffic will tend to avoid it relative to the pre-intervention case. In practice, of course, the effects may be more complex given the heterogeneity of movements: through-trips may avoid passing through E, which may release capacity that attracts shorter trips, but the observations show increases in both major through movements EA & (especially) EK. The network equilibrium model could never replicate this, and in fact would predict an effect in the opposite direction, although in this case it does show that the effects on total volume passing E are very small. Given the remarks made in §4 on the substantial level of ambient daily variability in E, what we are seeing in the observed data could well be explicable by such ambient variation, manifested now in the observed 'sample mean' flows. Whatever the source of this effect, it will confound any comparison between modelled and observed data; nevertheless, even though the modelled flows for EA and EK during the intervention are rather low, the model does predict the understandable rise of EK to become the main through-movement observed as EA falls. However, the fact that the absolute levels of EA and EK in the model are too low could be speculated to be indicative of behavioural features that cannot be captured in the Wardrop equilibrium model. These could include a habitual tendency to stay on familiar routes, inertia caused by the difficulty in predicting travel times in such a variable situation or by drivers adopting satisficing decision rules, or a transient effect (as travellers learn of the new traffic conditions over days) that may have decayed had the capacity reduction been in place for longer. While models exist in the literature to reflect any of these behavioural features and which could no doubt be fitted to the data observed, it is difficult to identify without additional evidence which of these behavioural paradigms would be most suitable in truly describing the underlying causality.

6. CONCLUSIONS

In the paper we have reported the findings of an empirically-driven investigation of the network effects of a road capacity reduction, performing statistical analyses of the results to account for ambient daily variation, and then we have compared these findings with those predicted by a network equilibrium model. A particular feature of the partial registration plate data used is that it can provide evidence of site-to-site movements (i.e. not just point surveillance data) in a non-intrusive way, and we have shown how careful analysis can remove concerns about matching the resulting data. Such data is potentially considerably richer in information terms than a much greater quantity of point-data. In the empirical analysis, the impact of the capacity change on the location, dispersion and shape of the travel time probability distribution was evident. Considering observations across days, it was seen how considerable ambient daily variation in flows could easily mask any systematic effect, but that this effect could be partially avoided by considering flow proportions between sites, providing statistically significant evidence of changes in routing patterns as a result of the interventions modelled.

The comparison of the observed data with the base (pre-intervention) network equilibrium model outputs revealed a major influence of a particular parameter that typically attracts little attention in network analysis: namely that parameter controlling the relative valuation of travel distance in the generalised cost equation. With a calibration of this parameter, the model was seen to provide a reasonable estimate of the site-to-site flows and travel times predicted in the vicinity of the intervention site, though the parameter value that achieved this fit was substantially lower than that typically obtained from economic studies in which distance is assumed a proxy for vehicle operating cost. It was demonstrated how such a lower value has an important spatial effect, placing greater emphasis on distance in the generalised cost equation, and thereby balancing flows towards radial routes at the expense of orbital routes, as was evident in the observed data. In predicting the impact of a capacity reduction, the network equilibrium model was more mixed in its performance. If a frictional effect of the capacity reduction is assumed and 'tuned', then the model could be made to give broad agreement with the direction of change observed in the data. On the other hand, there were features observed in the empirical data that could not be captured in the Wardrop equilibrium approach, however specified, with considerable ambient daily variation in flows leading to apparent paradoxical increases in demand for facilities whose capacity had been reduced, and with apparent inertia/learning effects in travellers' decisions.

The evidence in this study relates to two studies of real-life network interventions, and it would clearly be beneficial to conduct and report further studies in order to explore the generality and transferability of findings. This is probably more suited for a research study, where we can be more open about short-comings and negative findings, as opposed to practical studies for which there might be considerable political fall-out of any ambiguities in the model results. However, we feel that the low-cost, non-intrusive methodology used here could be adopted by local agencies wishing to periodically update and improve their transport network models, especially as the site-to-site flows and travel times can provide richer information for calibration than can surveillance data at a single point. This has been demonstrated in the present paper with the focus on the major spatial effects of the assumed valuations of travel time and distance in the generalised cost expression. Such a methodology would involve an agency:

a) Identifying in advance a planned network disruption, e.g. due to road or utilities maintenance, that is likely to have a significant impact on congestion and route choice.
b) Designing a registration plate survey of the kind described in section 2 to monitor routing patterns before, during and after the disruption.
c) Comparing the before-and-after data with the predictions of a network equilibrium model, adjusting the network model so as to reflect the empirical findings.
d) Periodically repeating steps a)–c), reflecting additionally on the consistency of findings from step c) over time.

In future research, the range of modelling methods considered may also be expanded, for example to include network models incorporating notions of day-to-day variability, within-day dynamic effects and departure time choice (e.g. Xie & Olszewski, 2011), or behavioural laws of learning/habit (see, for example, Guo & Liu, 2011; Gao et al, 2011). On the other hand, we believe it also relevant to test simpler models (such as the static Wardrop equilibrium model considered in the present paper) to be clear on the benefits of more complex specifications and/or those with higher numbers of parameters. For example, we may propose a model with a time-sliced OD matrix, but can we estimate such a matrix to a reasonable degree of accuracy, and how stable are such estimates over time? In the end, would this give us a more accuracy picture of traffic conditions, or just a model with more parameters that might be poorer for prediction? Finally, as emerging data sources become more widely available, such as those based on mobile communications traces, we will have the potential (in principle at least) to more accurately link data spatially across a network, and even to connect the trajectories followed by individuals on different days. This has the potential to provide the richest kind of data for empirical analysis of network

models, yet still many of the conceptual and methodological issues raised in the present paper will be pertinent.

ACKNOWLEDGEMENTS

REFERENCES

Abu-Eisheh SA & Mannering FL (1987). Discrete/continuous analysis of commuters' route and departure time choices. Transportation Research Record **1138**, 27-34.

Antonisse RW, Daly AJ & Ben-Akiva M (1989). Highway assignment method based on behavioural models of car drivers' route choice. Transportation Research Record **1220**, 1-11.

Bar-Gera, H (2010). Traffic assignment by paired alternative segments. Transportation Research **44B**, 1022–1046.

BBC (2000a). UK fuel shortages worsen. BBC news, http://news.bbc.co.uk/1/hi/uk/919429.stm, visited March 5th 2009.

BBC (2000b). Flooding in recent years. BBC news, http://www.bbc.co.uk/weather/features/flooding_past.shtml

Beckmann M, McGuire C & Winston C (1956). Studies in the economics of transportation. Yale University Press, New Haven, CT.

Bekhor S, Ben-Akiva ME & Ramming MS (2002). Adaptation of logit kernel to route choice situation. Transportation Research Record **1805**, 78-85.

Ben-Akiva M, Davol A & Toledo T (2000). MITSIM: Enhancements, Calibration & Validation. Research Report, Intelligent Transport Systems Program, Massachusetts Institute of Technology.

Berdica K (2002). An introduction to road vulnerability: what has been done, is done and should be done. Transport Policy **9**(2), 117-127.

Bonsall PW, Montgormery FO & Jones C (1984). Deriving the Constancy of Traffic Flow Composition from Vehicle Registration Data. Traffic Engineering & Control **25**(7/8), 386-391.

Chen T-Y, Chang H-L & Tzeng G-H (2001). Using a weight-assessing model to identify route choice criteria and information effects. Transportation Research **35A**(3), 197-224.

DfT (1989). Highways Economics Note No.2 in COBA 9 Manual. Department of Transport, HMSO, London.

Dial, RB (2006). A path-based user-equilibrium traffic assignment algorithms that obviates path storage and enumeration. Transportation Research **40B**(10), 917-936.

Duffell JR & Kalombaris A (1988). Empirical studies of car driver route choice in Hertfordshire. Traffic Engineering & Control **29**(7/8), 398-408.

Emmerink RHM, Nijkamp P, Rietveld P & Van Ommeren JN (1996). Variable message signs and radio traffic information, an integrated empirical analysis of drivers' route choice behaviour. Transportation Research **30A**(2), 135-153.

Florian M & Nguyen S (1976). An application and validation of equilibrium trip assignment methods. Transportation Science **10**(4), 374-390.

Fox K, Clark S, Boddy R, Montgomery F & Bell M (1998). Some benefits of a SCOOT UTC system. Traffic Engineering & Control **39**(9), 484-489.

Gao S, Frejinger E & Ben-Akiva M (2011). Cognitive cost in route choice with real-time information: An exploratory analysis. Transportation Research A, in press.

Goodwin P (1998). The End of Equilibrium. In The Theoretical Foundations of Travel Choice Modelling (eds: Gärling T, Laitila T & Westin K), 103-127, Pergamon, Oxford.

Goulias KG (1999). Longitudinal analysis of activity and travel pattern dynamics using generalized mixed Markov latent class models. Transportation Research **33B**(8), 535-558.

Guo X & Liu HX (2011). Bounded rationality and irreversible network change. Transportation Research B, in press.

Highways Agency (2002). Traffic Appraisal of Road Schemes. Design Manual for Roads & Bridges,Vol 12.http://www.archive.official-documents.co.uk/document/ha/dmrb/index.htm

Huchingson R, McNees R & Dudek C (1977). Survey of motorist route-selection criteria. Transportation Research Record **643**, 45-48.

Huff JO & Hanson S (1986). Repetition and variability in urban travel. Geographical Analysis **18**(2), 97-114.

Iida Y, Aikyama T & Takashi U (1992). Experimental analysis of dynamic route choice behaviour. Transportation Research **26B**(1), 17-32.

Iida Y (1999). Basic concepts and future directions of road network reliability analysis. Journal of Advanced Transportation **33**(2), 125-134.

Jan O, Horowitz AJ & Peng ZR (2000). Using global positioning system data to understand variations in path choice. Transportation Research Record **1725**, 37-44.

Janson BN (1991). Dynamic traffic assignment for urban road networks. Transportation Research **25B**(2/3), 143-161.

Janson BN, Thint SPT & Hendrickson CT (1986). Validation and use of equilibrium network assignment for urban highway reconstruction planning. Transportation Research **20A**(1), 61-73.

Jou R-C (2001). Modeling the impact of pre-trip information on commuter departure time and route choice. Transportation Research **35B**(10), 887-902.

Keenan D (2005). M62-M1 interchange – a combined TRANSYT/VISSIM micro-simulation assessment. Traffic Engineering and Control 46(5), 178-187.

Khattak AJ, Koppelman FS & Schofer JL (1993). Stated preferences for investigating commuters' diversion propensity. Transportation **20**(2), 107-127.

Laird J, Druitt S & Frasier D (1999). Edinburgh city centre: A micro-simulation case study. Traffic Engineering & Control **40**(2) 72-76.

Leonard DR & Tough JB (1979). Validation work on CONTRAM - a model for use in the design of traffic management schemes. Proceedings PTRC Summer Annual Meeting, Traffic & Environmental Management Seminar, 135-153, PTRC Ltd., London.

Liu HX, Recker W & Chen A (2004). Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data. Transportation Research **38A**, 435–453.

Lyons G & Chaterjee K (2002). Transport Lessons from the Fuel Tax Protests of 2000. Ashgate, UK.

Mahmassani HS & Herman R (1990). Interactive experiments for the study of tripmakers behaviour dynamics in congested commuting systems. In: Jones P (ed), Developments in Dynamic and Activity-Based Approaches to Travel Analysis, Chapter 13, Avebury, Aldershot, UK.

Mahmassani HS & Jou R-C (2000). Transferring insights into commuter behavior dynamics from laboratory experiments to field trials. Transportation Research **34A**(4), 243-260.

Mannering F, Kim S-G, Barfield W & Ng L (1994). Statistical analysis of commuters' route, mode, and departure time flexibility. Transportation Research **2C**(1), 35-48.

Matzoros T, Randle J, Van Vliet D & Weston B (1987). A validation of the SATURN and ME2 models using before-and-after survey data from Manchester. Traffic Engineering & Control **28**(12), 641-643.

May AD, Bonsall PW & Marler NW (1989). Car travel time variability on links of a radial route in London. Working Paper 279, Institute for Transport Studies, University of Leeds.

May AD & Milne DS (2000). Effects of alternative road pricing systems on network performance. Transportation Research **34A**(6), 407–435.

McLeod FN, Hounsell NB & Bain RS (1993). Highway reliability supply effects. Proceedings PTRC Summer Annual Meeting, Seminar D, 191-202, PTRC Ltd.

Montgomery FO & May AD (1987). Factors affecting travel times on urban radial routes. Traffic Engineering & Control **28**(9), 452-458.

Nakayama S, Connors R, Watling D (2009). Estimation of Parameters of Network Equilibrium Models: A Maximum Likelihood Method and Statistical Properties of Network Flow. In: Proc International Symposium on Transportation & Traffic Theory 2009 (eds: WHK Lam, SC Wong & HK Lo), Springer, 39-56.

Nie Y (2010). A class of bush-based algorithms for the traffic assignment problem. Transportation Research **44B**(1), 73-89.

Noland RB (2001). Relationships between highway capacity and induced vehicle travel. Transportation Research **35A**(1), 47-72.

Quiroga C, Henk R & Jacobson M (2000). Innovative data collection techniques for roadside origin-destination surveys. Transportation Research Record **1719**, 140-146.

Rakha H & Van Aerde M (1995). Statistical analysis of day-to-day variations in real-time traffic flow data. Transportation Research Record **1510**, 26-34.

Ribeiro PCM (1994). Handling traffic fluctuation with fixed-time plans calculated by TRANSYT. Traffic Engineering & Control **35**(6), 362-366.

Schaefer MC (1988). License Plate Matching Surveys: Practical Issues and Statistical Considerations. ITE Journal **58**(7), 37-42.

Smeed RJ & Jeffcoate GO (1971). The variability of car journey times on a particular route. Traffic Engineering & Control **13**, 238-243.

Stephenson B & Teply S (1984). The verification of CONTRAM in Edmonton. Traffic Engineering & Control **25**(7), 376-385.

Sumalee A, May AD & Shepherd SP (2005). Comparison of judgemental and optimal road pricing cordons. Transport Policy **12**, 384-390.

Taylor MAP, Woolley JE & Zito R (2000). Integration of the Global Positioning System and Geographical Information Systems for Traffic Congestion Studies. Transportation Research **8C**, 257-285.

Taylor MAP, Sekhar SVC & D'Este GM (2006). Application of accessibility based methods for vulnerability analysis of strategic road networks. Networks & Spatial Economics **6**, 267-291.

TMIP (1997). Model Validation and Reasonableness Checking Manual. Chapter 7, US Dept of Transportation http://tmip.fhwa.dot.gov/clearinghouse/docs/mvrcm/ch7.stm .

Van Vliet D (1982). SATURN − a modern assignment model. Traffic Engineering & Control **23**(12), 578–581.

Vaughn KM, Reddy P, Abdel Aty MA, Kitamura R & Jovanis P (1995). Route choice and information use: initial results from simulation experiments. Transportation Research Record **1516**, 61-69.

Wachs M (1967). Relationships between drivers' attitudes toward alternative routes and driver and route characteristics. Highway Research Record **197,** 70-87.

Wardman MR, Bonsall PW & Shires JD (1997). Stated preference analysis of driver route choice reaction to variable message sign information. Transportation Research **5C**(6), 389-405.

Watling DP (1990). The statistical analysis of partial registration plate data. PhD thesis, Department of Probability & Statistics, University of Sheffield, UK.

Watling DP (1994). Maximum likelihood estimation of an origin-destination matrix from a partial registration plate survey. Transportation Research **28B**(4), 289-314.

Watling DP & Maher MJ (1988). A graphical procedure for analysing partial registration plate data. Traffic Engineering & Control **29**(10), 515-519.

Xie L & Olszewski P (2011). Modelling the effects of road pricing on traffic using ERP traffic data. Transportation Research **45A**, 512–522.

Xu G, Lam WHK & Chan KS (2004). Integrated approach for Trip Matrix Updating and Network Calibration. Journal of Transportation Engineering **130**(2), 231-244.

Yang H, Iida Y & Sasaki T (1994). The equilibrium-based origin-destination matrix estimation problem. Transportation Research **28B**(1), 23-33.

Zhu S, Levinson D, Liu HX & Harder K (2010). The traffic and behavioral effects of the I-35W Mississippi River bridge collapse. Transportation Research **44A**, 771-784.

Table 1: Fishergate study, maximum likelihood estimates for journeys emanating from site E: 'during intervention' days shaded (full survey period at site E)

Date	Flows					Journey time mean				Journey time standard dev.			
	EA	EB	EK	EF	E total	EA	EB	EK	EF	EA	EB	EK	EF
26/6	289	18	237	54	1387	8.96	9.95	8.49	7.29	1.75	2.25	1.43	1.18
27/6	270	19	229	48	1306	9.29	9.51	9.00	7.23	2.04	1.56	2.54	1.38
28/6	310	13	289	54	1469	7.90	8.39	5.22	7.26	2.92	1.68	1.71	1.73
29/6	300	22	283	68	1448	5.47	6.08	5.05	7.24	1.20	2.12	1.11	1.42
2/7	319	22	291	70	1527	6.21	5.97	5.55	6.95	1.61	1.75	1.08	1.05
3/7	234	11	325	61	1448	12.12	9.38	6.72	6.78	7.96	4.83	3.27	0.96
4/7	234	12	328	66	1494	11.52	10.08	6.26	7.20	5.69	3.06	2.10	1.54
5/7	240	13	358	77	1522	11.06	11.23	4.95	7.81	4.96	6.61	0.92	3.14
6/7	265	17	352	69	1518	8.48	6.93	4.84	6.72	4.27	3.75	1.02	1.34
11/7	261	12	272	78	1481	8.67	11.97	4.98	6.87	5.12	6.42	1.33	1.62
12/7	252	13	329	67	1512	9.84	11.70	4.93	6.83	4.78	7.60	1.21	0.77
13/7	274	14	318	103	1603	5.14	6.22	4.58	7.55	1.02	3.05	0.94	1.69
16/7	286	17	357	65	1578	9.50	9.17	5.21	6.32	5.51	4.93	1.32	0.76

Table 2: Fishergate study, maximum likelihood estimates for E to A journeys on 26/6 (before intervention): impact of truncation of survey period at site E

Time span at E	Mean	St. dev	alpha	Matched flow
7:45 to 8:35	8.74	1.72	0.2321	186
7:45 to 8:40	8.83	1.77	0.2294	199
7:45 to 8:45	9.00	1.81	0.2285	213
7:45 to 8:50	9.10	1.80	0.2275	228
7:45 to 8:55	9.18	1.74	0.2290	243
7:45 to 9:00	9.13	1.70	0.2313	262
7:45 to 9:05	9.01	1.76	0.2317	281
7:45 to 9:10	8.94	1.75	0.2235	290
7:45 to 9:15	8.94	1.75	0.2095	290

Table 3: Fishergate study, maximum likelihood estimates for E to A journeys on 3/7 (during intervention): impact of truncation of survey period at site E

Time span at E	Mean	St. dev	alpha	Matched flow
7:45 to 8:35	9.53	5.86	0.2018	186
7:45 to 8:40	10.73	6.91	0.2029	204
7:45 to 8:45	11.96	7.88	0.2002	222
7:45 to 8:50	12.52	8.17	0.2000	234
7:45 to 8:55	12.52	8.08	0.1932	235
7:45 to 9:00	12.42	8.12	0.1820	237
7:45 to 9:05	12.32	8.06	0.1680	238
7:45 to 9:10	12.26	7.99	0.1543	235
7:45 to 9:15	12.17	8.01	0.1493	234

Table 4: Fishergate study, maximum likelihood estimates for journeys emanating from site E: 'during intervention' days shaded (with truncated survey period at site E)

Date	Flows					Journey time mean				Journey time standard dev.			
	EA	EB	EK	EF	E total	EA	EB	EK	EF	EA	EB	EK	EF
26/6	243	15	205	50	1044	9.18	10.53	8.60	7.43	1.74	1.95	1.32	1.16
27/6	234	15	215	49	966	9.38	9.51	9.05	8.15	2.17	1.67	2.61	2.61
28/6	258	13	251	50	1131	8.32	8.50	5.36	7.30	3.03	1.64	1.83	1.83
29/6	257	19	235	59	1161	5.61	6.67	5.24	7.28	1.08	2.10	0.85	1.37
2/7	258	19	259	51	1195	5.96	5.62	5.46	7.14	1.63	1.57	1.06	1.12
3/7	235	11	303	52	1134	12.52	9.47	6.43	6.82	8.08	4.88	3.19	1.01
4/7	231	12	282	56	1153	11.80	10.13	5.85	7.39	5.62	3.10	1.90	1.84
5/7	230	13	295	67	1174	10.79	10.49	4.84	7.05	5.05	7.52	0.93	2.01
6/7	253	13	294	59	1188	8.22	6.12	4.73	6.91	4.06	2.22	1.04	1.29
11/7	231	10	266	64	1113	8.31	12.89	5.01	7.57	4.90	7.91	1.35	2.34
12/7	244	13	275	52	1223	9.72	11.66	5.00	6.94	4.69	8.06	1.33	0.76
13/7	226	14	273	96	1265	5.21	6.59	4.60	7.63	1.05	3.33	0.96	1.78
16/7	272	17	288	58	1246	9.33	9.26	4.97	6.40	5.37	4.94	1.26	0.77

Table 5: Maximum likelihood estimates of matched vehicles emanating from site A crossing bridges in the Lendal Bridge study, 'during intervention' day shaded

Date	Total flow at Blossom St (site A)	Matched flow from site A to …		
		Lendal (site H)	Ouse (site J)	Skeldergate (site C)
27/6	843	157	58	165
28/6	880	149	120	192
6/9	735	137	102	
7/9	635	129	91	136
8/9	606	115	87	94
27/9	546		118	141

Table 6: Percentage of generalised cost attributable to (weighted) travel time at different speeds and different relative valuations of time and distance

Speed (km/h)	Assumed valuation ratio α/β (km/minute)					
	2	**1**	**0.5**	**0.3**	**0.2**	**0.1**
10	92	86	75	64	55	38
20	86	75	60	47	38	23
30	80	67	50	38	29	17
40	75	60	43	31	23	13
50	71	55	38	26	19	11
60	67	50	33	23	17	9
80	60	43	27	18	13	7
100	55	38	23	15	11	6

Table 7: Comparison of observed data and modelled data for Fishergate study, in base scenario (without capacity reduction, 13/7 day neglected)

Scenario	Flows					Journey time mean			
	EA	EB	EK	EF	E total	EA	EB	EK	EF
Observed	250	16	233	52	1099	7.7	8.2	6.7	7.5
Modelled (α/β = 2)	104	0	45	1	880	5.3	5.6	4.5	5.2
Modelled (α/β = 1)	146	0	95	5	995	5.8	6.2	5.0	5.4
Modelled (α/β = 0.5)	148	0	139	36	1107	6.4	6.7	5.5	5.8
Modelled (α/β = 0.3)	187	0	156	64	1204	8.1	8.4	7.2	6.0
Modelled (α/β = 0.2)	229	0	150	66	1250	8.2	8.5	7.5	6.7
Modelled (α/β = 0.1)	296	0	176	61	1400	12.0	12.2	11.0	6.3
Modelled (α/β = 0)	225	0	411	0	1387	16.9	16.2	16.2	5.2

Table 8: Comparison of observed data and modelled data for Fishergate study ($\alpha/\beta = 0.2$), before and during capacity reduction, under different Saturation Flow (SF) assumptions (observed data for 13/7 and 16/7 days neglected)

Scenario	Flows					Journey time mean			
	EA	EB	EK	EF	E total	EA	EB	EK	EF
Observed (full capacity)	250	16	233	52	1099	7.7	8.2	6.7	7.5
Modelled (full capacity)	229	0	150	66	1250	8.2	8.5	7.5	6.7
Observed (reduced capacity)	237	12	286	58	1164	10.2	10.1	5.3	7.1
Modelled (SF = 50% of base)	234	0	150	66	1253	8.3	8.5	7.5	6.6
Modelled (SF = 40% of base)	171	0	172	64	1229	10.1	9.8	6.2	6.6
Modelled (SF = 33% of base)	128	0	200	63	1226	12.6	9.0	5.8	6.3
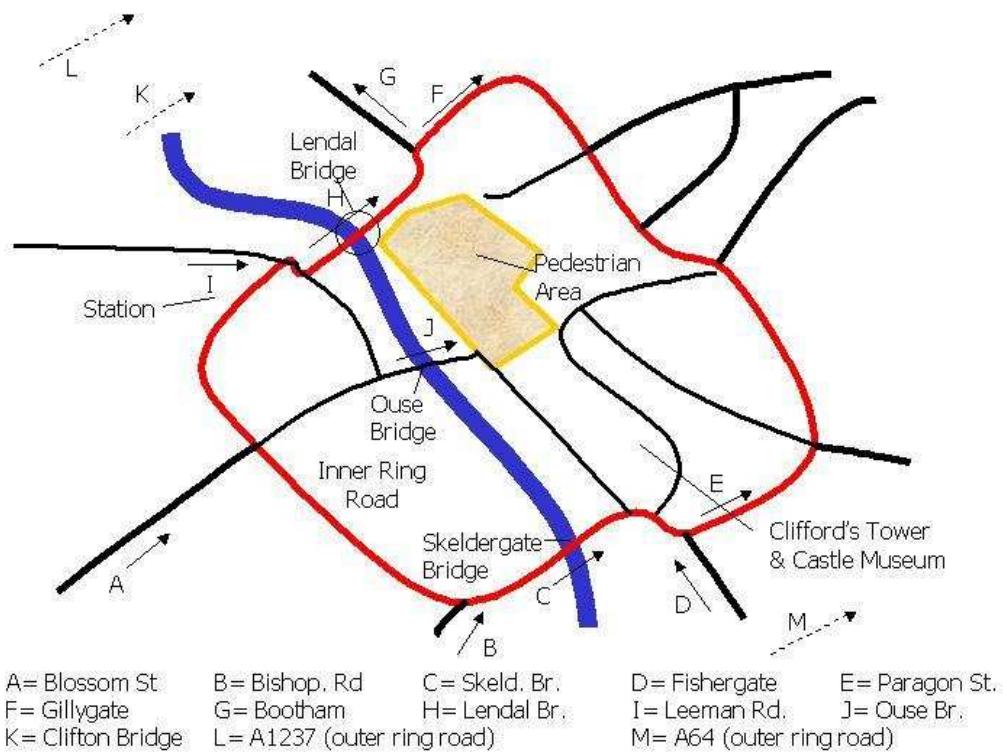
Figure 1: Intervention and survey locations for Lendal Bridge study

A= Blossom St B= Bishop. Rd C= Skeld. Br. D= Fishergate E= Paragon St.
F= Gillygate   G= Bootham   H= Lendal Br. I= Leeman Rd. J= Ouse Br.
K= Clifton Bridge L= A1237 (outer ring road) M= A64 (outer ring road)

A= Fishergate (left)
B= Fishergate (right)
C= Paragon St.
D= Skeldergate Br.
E= Fulford Rd.
F= Uni. Rd.
G= Lawrence St.
H= Blossom St.
I= Queen St.
J= Rougier St.
K= Cemetery Rd.
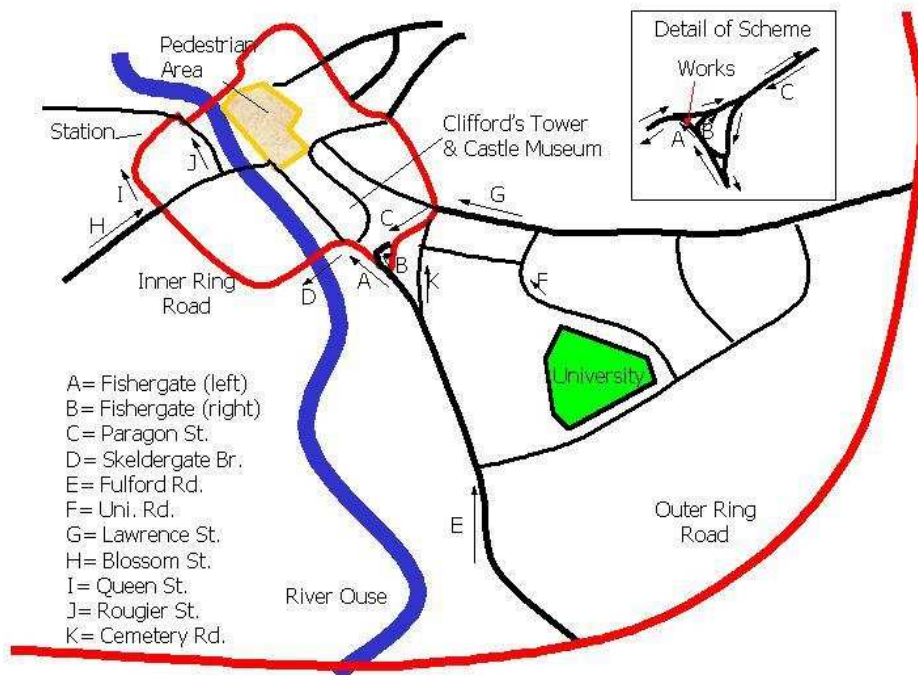
Figure 2: Intervention and survey locations for Fishergate study

33
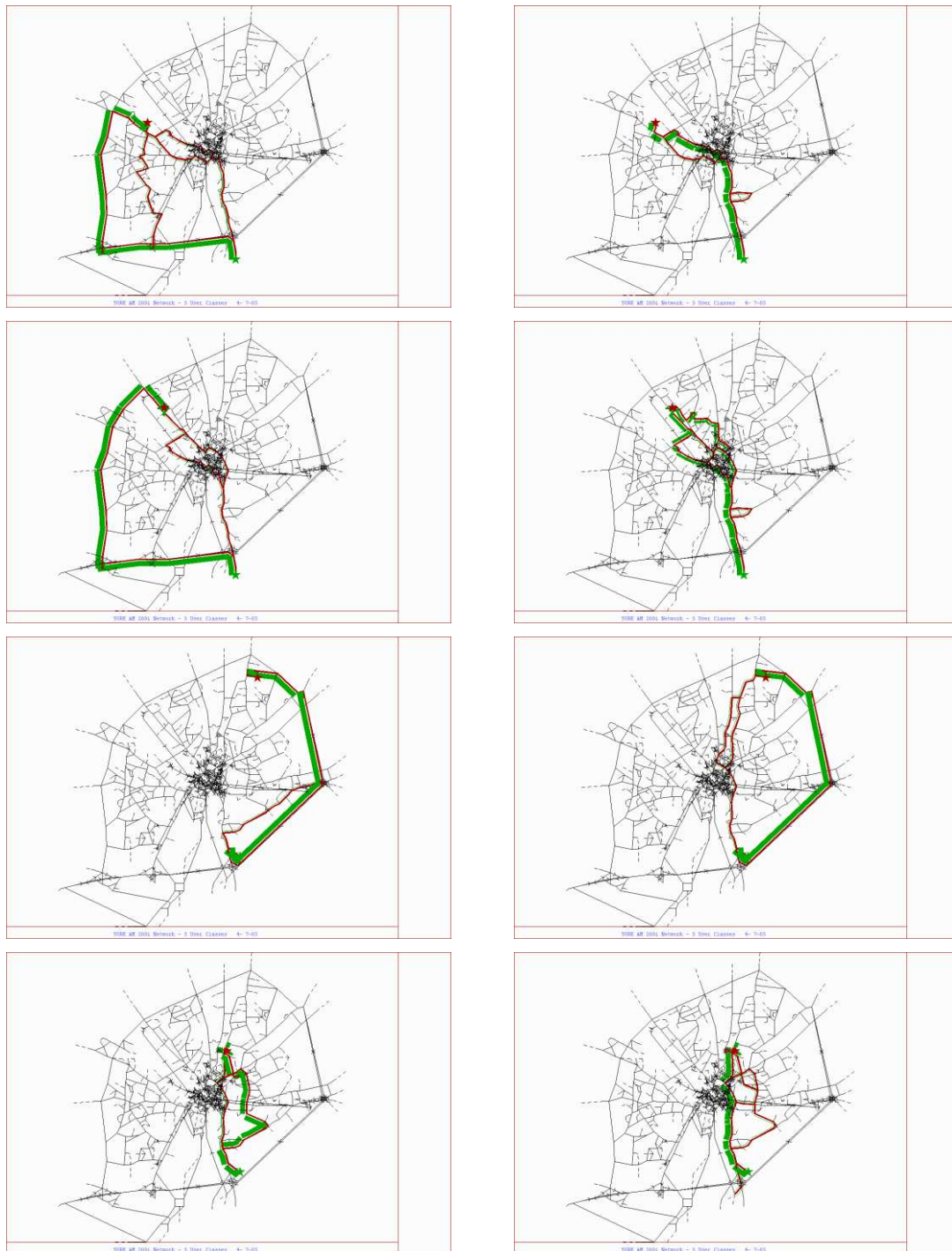
Initial Route Choices (α/β = 2)   **Calibrated Route Choices (α/β = 0.2)**

**Figure 7: Comparison of selected initial and 'calibrated' route choices
based on network equilibrium model for Fishergate base scenario**
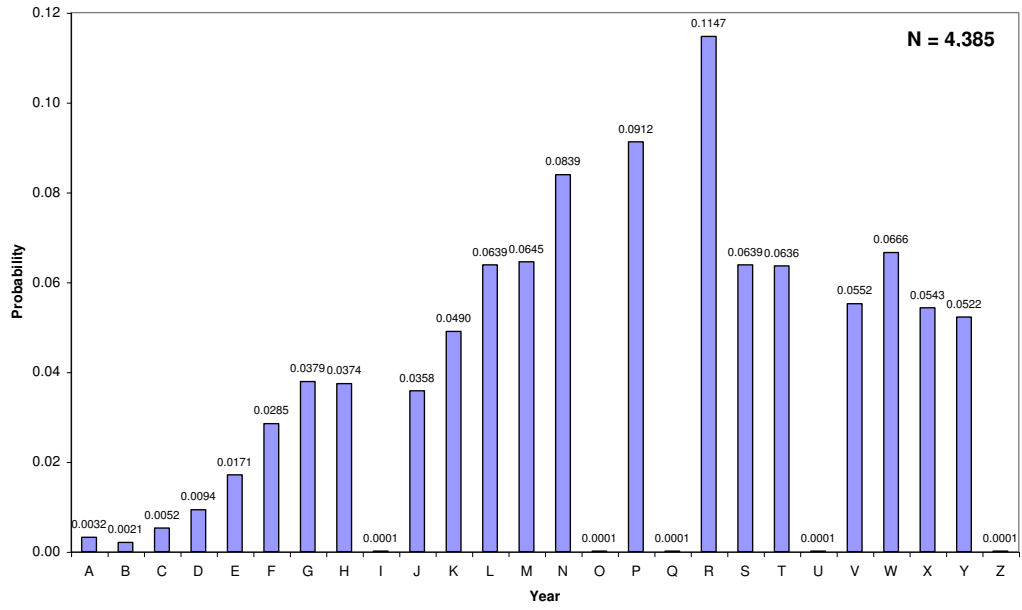
Figure 3: Empirical distribution of year letters for Fishergate study
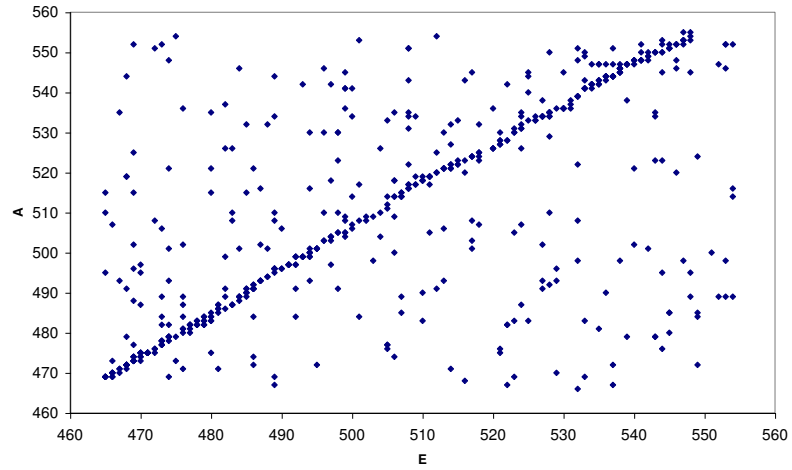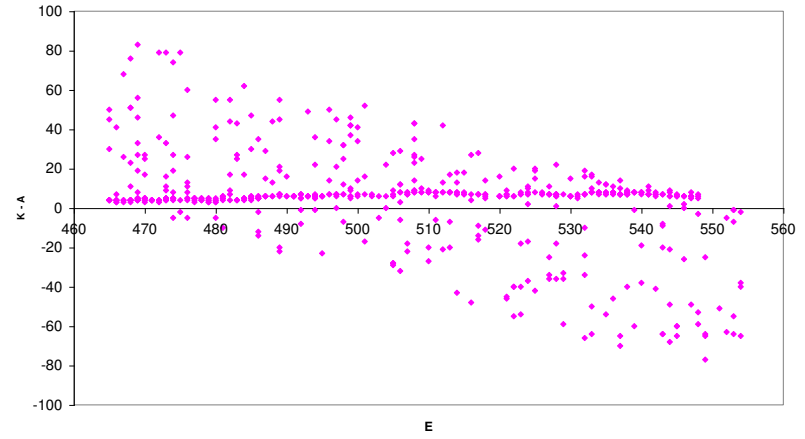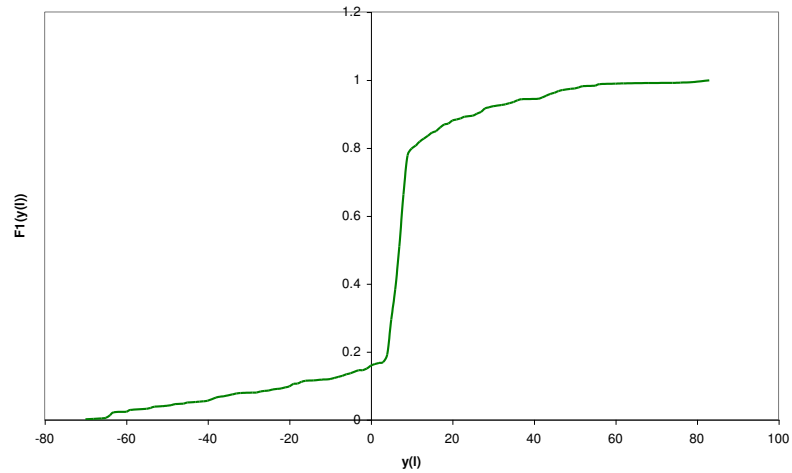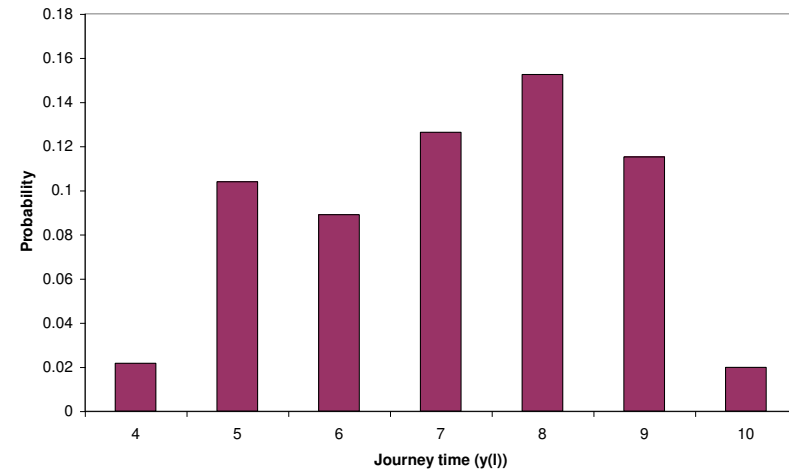
Figure 4: Matched partial plates between sites E & A, Fishergate study, July 2nd (before intervention):
(a) scatter plot of matched time-stamps at E and A; (b) implied journey times of partial plates against time-stamp at E;
(c) estimated cumulative frequency distribution for journey times of genuine matches;
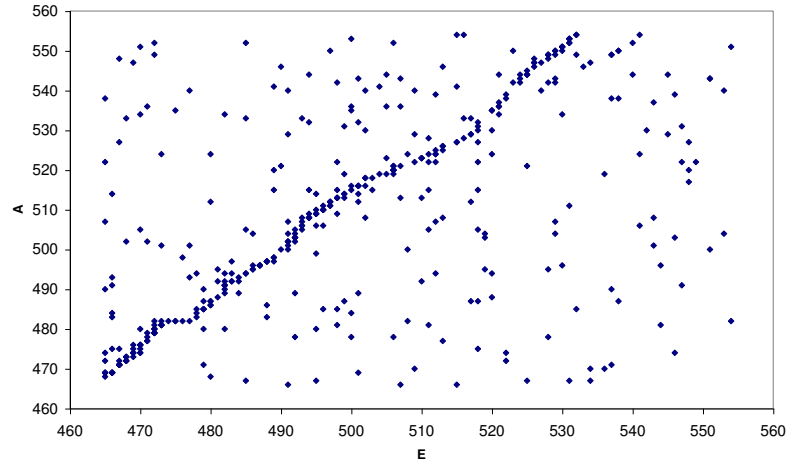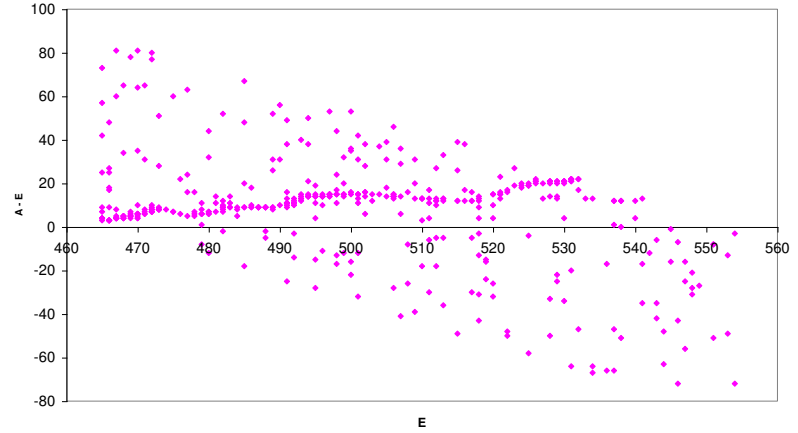**(d) estimated genuine journey time probability distribution.**
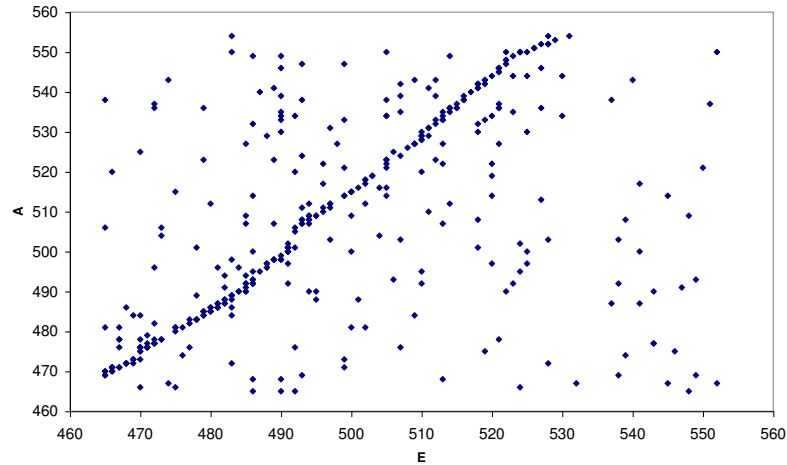
(a)
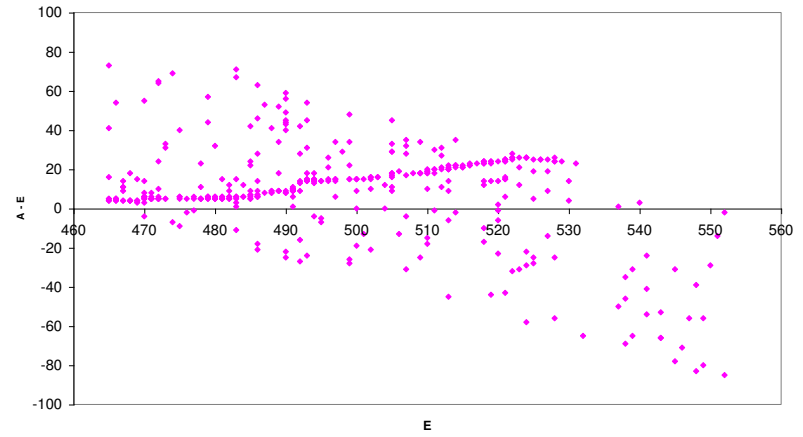
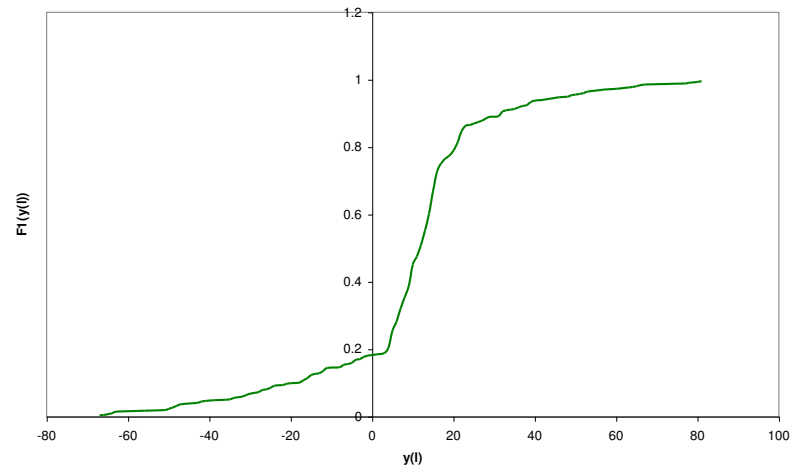**(b)**
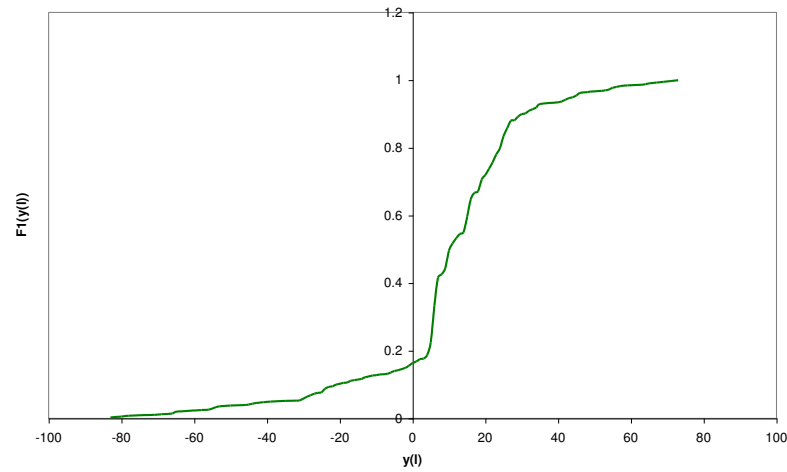
(c)
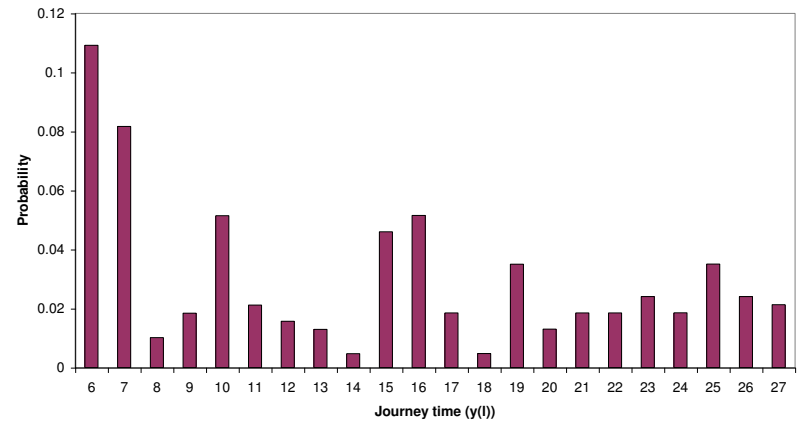
(d)

Figure 5: Matched partial plates between sites E & A, Fishergate study, July 3rd (after intervention):
(a) scatter plot of matched time-stamps at E and A; (b) implied journey times of partial plates against time-stamp at E;
(c) estimated cumulative frequency distribution for journey times of genuine matches;
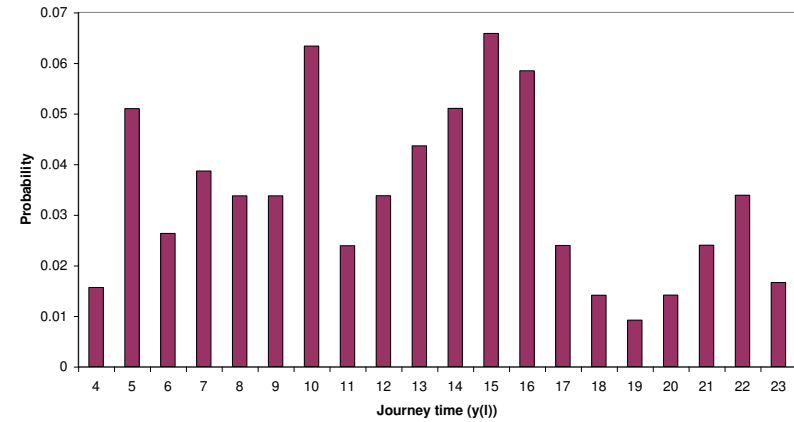**(d) estimated genuine journey time probability distribution.**

Figure 6: Matched partial plates between sites E & A, Fishergate study, July 4th (after intervention):
(a) scatter plot of matched time-stamps at E and A; (b) implied journey times of partial plates against time-stamp at E;
(c) estimated cumulative frequency distribution for journey times of genuine matches;
(d) estimated genuine journey time probability distribution.