Wilson, R. E., & Lunt, G. (2002). New data sets and improved models of highway traffic.

New Data Sets and Improved Models of Highway Traffic

George Lunt, TRL Ltd. glunt@trl.co.uk
R. Eddie Wilson[1], University of Bristol. RE.Wilson@bristol.ac.uk

Abstract

In this paper we report progress on the use of novel data sets to calibrate and improve car-following models of highway traffic. These models describe vehicles as discrete entities moving in continuous time and space, and use plausible behavioural assumptions to derive systems of differential equations which must be solved for vehicles' trajectories. To date, there has been little or no systematic attempt to fit these models to quantative microscopic (at the level of individual vehicles) highway data. Our recent work has involved the analysis of two such data sets:

- MIDAS (Motorway Incident Detection and Automatic Signalling) IVD (Individual Vehicle Data), and
- aerial video pictures.

Each of these data sets has been kindly provided by the Highways Agency.

The MIDAS (Motorway Incident Detection and Automatic Signalling) system consists of double inductance loops which record the lane number, time headway and wall-clock time of passing vehicles, as well as estimating their lengths and speeds. This data has been collected by TRL.

We have applied novel statistical techniques to the pattern matching of vehicle length measurements at different loops, and thus we are able to track individual vehicles down the motorway. In particular, we may:

- obtain accurate estimates for individual vehicles' travel times between loops;
- estimate lane changing rates.

However, presently MIDAS loops may not be used to estimate vehicles' accelerations. We describe how image analysis and pattern matching techniques have been used to automatically follow vehicles across a sequence of aerial video frames. Projective transformations can then be applied to build full time series of vehicles' displacements, from which headway, velocity, and acceleration information can be extracted.

1 Introduction

In this paper we report progress on our programme of research dedicated to the improvement of highway traffic models. Broadly speaking these models fall into two categories:

1. Simplified models which have appeared in the applied mathematics and theoretical physics literatures since the mid 1990s. The principle goal of these models is to illustrate interesting phase change phenomena in an attempt to gain a qualitative understanding of traffic jams.
2. Models devised by engineers which are incorporated in commercial simulation software. Typically these models are very complex and have large numbers of parameters which purport to model driver behaviour and variability in detail.

To illustrate our division, consider so-called *car-following models* which describe vehicles as discrete entities moving in continuous time and space, see Fig. 1. In such models, acceleration laws are written down which attempt to model drivers' responses to external stimuli, which include such factors as the drivers' speed, and the relative displacement (i.e., headway) and relative velocities of neighbouring vehicles.

[1]Supported by the Nuffield Foundation awards for Newly Appointed Lecturers in Science, Engineering and Mathematics, grant number NAL/00544/G.

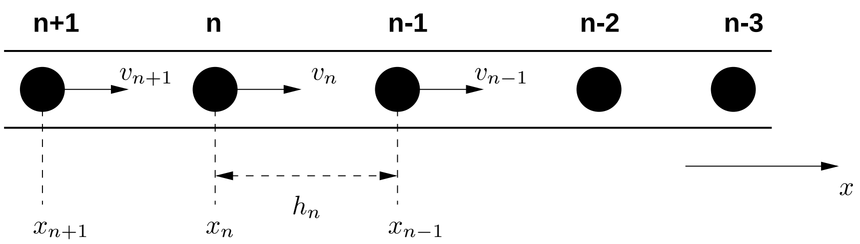

Figure 1: Common notation for car-following models. Here we show a unidirectional flow with one lane only and vehicles, which are labelled **1**, **2**, **3**, . . . etc. in an upstream direction, travel to the right in the increasing x direction. The displacements and velocities of vehicles are labelled $\dots$, x_{n-1}, x_n, x_{n+1}, $\dots$ and $\dots$, v_{n-1}, v_n, v_{n+1}, $\dots$ respectively. Of interest is the separation of consecutive vehicles, known as the *headway*. We define the headway of vehicle n to be $h_n = x_{n-1} - x_n$.

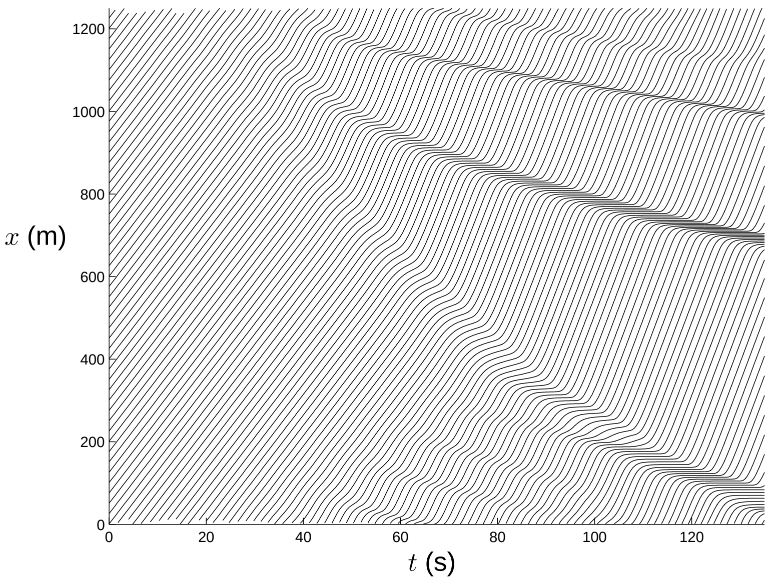

Figure 2: Illustration of instability and spontaneous formation of traffic jams for the optimal velocity model (1). Here the trajectories $x(t)$ are plotted when initial data is taken close to a uniform flow solution with separation $25\mathrm{m}$ and speed $\simeq 15\mathrm{ms}^{-1}$, and the sensitivity is taken below the stability threshold. The optimal velocity function V is taken from Bando et al (1995b).

A particularly simple acceleration law was suggested by Bando et al (1995a) and took the form

$$\dot{v}_n = \alpha \left[V(h_n) - v_n \right], \qquad \alpha > 0. \tag{1}$$

Here V is an *optimal velocity function*, with the properties (i) $V(h) = 0$ for $h < h_{\mathrm{crit}}$ (i.e., h_{crit} is the headway corresponding to the *jam density*); (ii) V is an increasing function; and (iii) $\lim_{h\to\infty} V(h) = V_\infty$, where V_∞ is (something like) the speed limit. The idea is that drivers would like to drive as fast as possible up to some limit V_∞, but in practice drive slower because of safety constraints, e.g., the distance to the vehicle ahead. Thus for a particular headway h, $V(h)$ gives the *optimal velocity*, i.e., the fastest safe speed. Thus the right hand side of (1) indicates that if travelling slower than the optimal velocity, $V(h_n) - v_n$ is positive and we accelerate; conversely, if travelling faster than the optimal velocity, we decelerate.

Equation (1) is clearly a gross simplification of driver behaviour and cannot possibly replicate the microscopic details of drivers' interactions. However, the major surprise is that particular parameter choices give emergent macroscopic dynamics which resemble closely (see Fig. 2) the well-known stop-and-go waves propagating upstream in motorway traffic.

Let us turn our attention to engineering simulation models, such as that derived by Gipps (1982), which forms the kernel of several commercial simulation packages. The basic hypothesis of Gipps's model is the following behavioural model:

> *What speed should I travel at now, given the behaviour of the vehicle in front one reaction time ago. If the vehicle in front comes to a stop at what I think is its hardest rate, and one reaction time later I commence braking at my hardest rate, I must come to a stop safely.*

This law may be formulated as a differential-difference equation. Wilson (2001) examined Gipps's model and performed a detailed analysis of the uniform flow solutions and their stability. In particular, that paper shows that uniform flow may only become unstable in unrealistic parameter regimes, and it follows that Gipps's model is not able to replicate stop-and-go waves. Further work at TRL by Abou Rahme and White (2001) simulated minor adaptations of Gipps's model in an attempt to produce stop-and-go waves, but found no satisfactory parameter regimes. In conclusion: although engineering modes make a great effort to capture the details of driver behaviour, they are often unable to replicate important features of macroscopic motorway dynamics.

Thus in our view there is a need to either validate existing highway traffic models, or to develop new ones, which capture the full complexity of individual driver behaviour, but which also yield the correct macroscopic (over several minutes and kilometres) dynamics. New sources of microscopic data, which incorporate acceleration information, will be required to fit these models.

This paper discusses progress with two new data sets:

1. MIDAS (Motorway Incident Detection and Automatic Signalling) IVD (Individual Vehicle Data). For some time the Highways Agency has stored MIDAS data in the form of one minute averages. This data is very useful for analysing bulk flow properties and macroscopic dynamics (e.g., see Dixon et al (2002) for the well known MTV plots which demonstrate stop-and-go waves), but cannot be used for fitting microscopic behavioural rules. However, in a recent project funded by the Highways Agency, TRL used extra equipment at MIDAS outstations to collect and store full unaveraged individual vehicle data. This data set includes details of the lane number, time headway and wall-clock time of every passing vehicle, and also contains estimates of their lengths and speeds.

2. Aerial video pictures. The Highways Agency has recently funded work to examine lane changing behaviour at motorway intersections which involved filming traffic from a helicopter. An ongoing project at the University of Bristol is considering the automatic image processing of these pictures, in order to obtain high resolution time series of the displacements of the many vehicles passing through the field of vision.

A number of entirely novel observations may be made from MIDAS IVD. For example, it is possible to examine the full variability of speed-headway relationships, and not just their bulk averages. One may also consider other interesting microscopic quantities such as the correlation of consecutive vehicles' velocities, and use this information to test statistical theories concerning platoons. All this data may be collected cheaply with no modification to the existing MIDAS infrastructure.

The principal disadvantages of MIDAS IVD are (i) it does not make measurements of vehicles' accelerations and (ii) it makes only point measurements in space. However, one of the main topics which we discuss here is the *vehicle re-identification* problem, where pattern matching techniques are applied to the estimates of vehicle lengths at consecutive loops in an attempt to track the progress of individual vehicles down a 1km section of motorway. In particular, this procedure might allow us (i) to calculate accurate travel times for individual vehicles and (ii) estimate the number of lane changes, by observing discrepancies between the vehicle length sequences at consecutive outstations. This process is extremely complex due to the inaccuracy of the vehicle length estimates. The details of our analysis follow in Section 2.

The chief advantage of aerial pictures is that acceleration information may be obtained by the double differentiation of the time series of vehicles' displacements. However these time series are noisy and the image processing algorithms presently require a great deal of manual intervention. A short announcement of our progress to date follows in Section 3.

2 MIDAS IVD analysis

The Highways Agency (HA) has developed an automatic speed-control environment known as *Controlled Motorways*. At its core is the Motorway Incident Detection and Automatic Signalling (MIDAS) system, which monitors current traffic conditions. Alerts are communicated to a central control computer, which in turn relays messages back to motorists via message signs. In particular, the system at present uses an algorithm known as HIOCC to set temporary (mandatory) speed limits upstream of queues, in an attempt to improve safety by smoothing traffic flow: as a result journey times and journey time reliability are also improved. In such conditions motorists are also advised not to change lanes, in order that the lane utilisation may be improved. The *Controlled Motorways* programme is being rolled out across the UK, but the focus of our analysis is on data collected from the south western section of the M25 where *Controlled Motorways* has been in operation since 1995.

The MIDAS detection system consists of outstations every 500m down the length of the motorway. Each outstation consists of a pair of inductance loops buried in the surface of the road and a smart signal processing and communications box. The magnetic field of passing vehicles induces a current in each loop and the system is wired so that the carriageway and lane number of the vehicle can be identified.

Further, for each passing vehicle, the times at which each half of the inductance loop pair is activated and deactivated are captured. The microcontroller then uses this information to give each vehicle a timestamp (from which, by comparison with the timestamp of the preceding vehicle, the front-to-front time headway may be calculated). Further, the difference in activation times and longitudinal separation of each half of the loop pair may be used to estimate the speed of each vehicle. The activation and deactivation times may also be used to calculate the *time over loop*, which combined with the speed estimate may be used to estimate the vehicle's length.

As we discuss further below, all measurements are rounded at various stages in the microcontroller's algorithm, and the nominal accuracy of recorded measurements are 1s, 0.1s, 1km/h and 1cm for the timestamp, time headway, speed and length respectively.

In the past, outstations have not stored the full details of individual vehicles, but have bundled the data into one minute averages which is then sent to the central control computer. However, TRL has now intercepted the individual vehicle data (IVD) before it is lost, and consequently we have access to comprehensive IVD sets from across the UK. This allows us to peform highway traffic analysis at a level of detail which has not previously been possible.

For example, it is now possible to analyse not only the bulk speed-density relation, but also the speed-headway information of individual vehicles: see Fig. 3(a). Thus we may analyse the huge variations in behaviour between different drivers.

A further avenue of research concerns the analysis of the correlation of speeds of consecutive vehicles, see Fig. 3(b). In particular, the skewness of these distributions, and how that skewness depends on lane number, may contain information about platoon behaviour.

The scatter plot of Fig. 3(b) indicates granularity in measured speeds which is examined further in the histogram of Fig. 3(c). It is clear that above 70km/h, speed measurements occur in narrow clusters, with almost no measurements between clusters. Further, the separation of clusters increases as speed increases. It is clear that the true accuracy of speed measurements, particularly of fast vehicles, is somewhat coarser than the quoted 1km/h. So far we have not had access to the details of the microcontroller algorithm, but these statistics could be explained by the use of rounded activation times in the speed calculation.

Unfortunately, the errors in speed measurements will have a knock-on effect on length measurements, whose errors might also be compounded by the use of rounded activation and deactivation times in the *time over loop* calculation. It is thefore possible for length measurements to be quite inaccurate.

We have performed some analysis of length measurement errors using existing IVD, in which the true lengths of vehicles are actually unknown. To do this, we examined the data sets obtained from two outstations (numbers 4897 and 4907) which are in the vicinity of junction 13 of the M25, between the busy interchanges with the M3 and M4. Outstations 4897 and 4907 are 1km apart, and the section of motorway between them is particularly simple, in that there are four continuous lanes in both directions and no on- or off-ramps[2]. Henceforth we restrict attention to the clockwise (northbound) carriageway and will refer to 4897 as the *upstream outstation* and 4907 as the *downstream outstation*.

Since there are no on- or off-ramps on this section of motorway, all vehicles which pass the upstream outstation must also pass the downstream outstation a short time later, and conversely,

[2]We have not considered outstation 4902 since its IVD appears to have a higher than usual error rate.

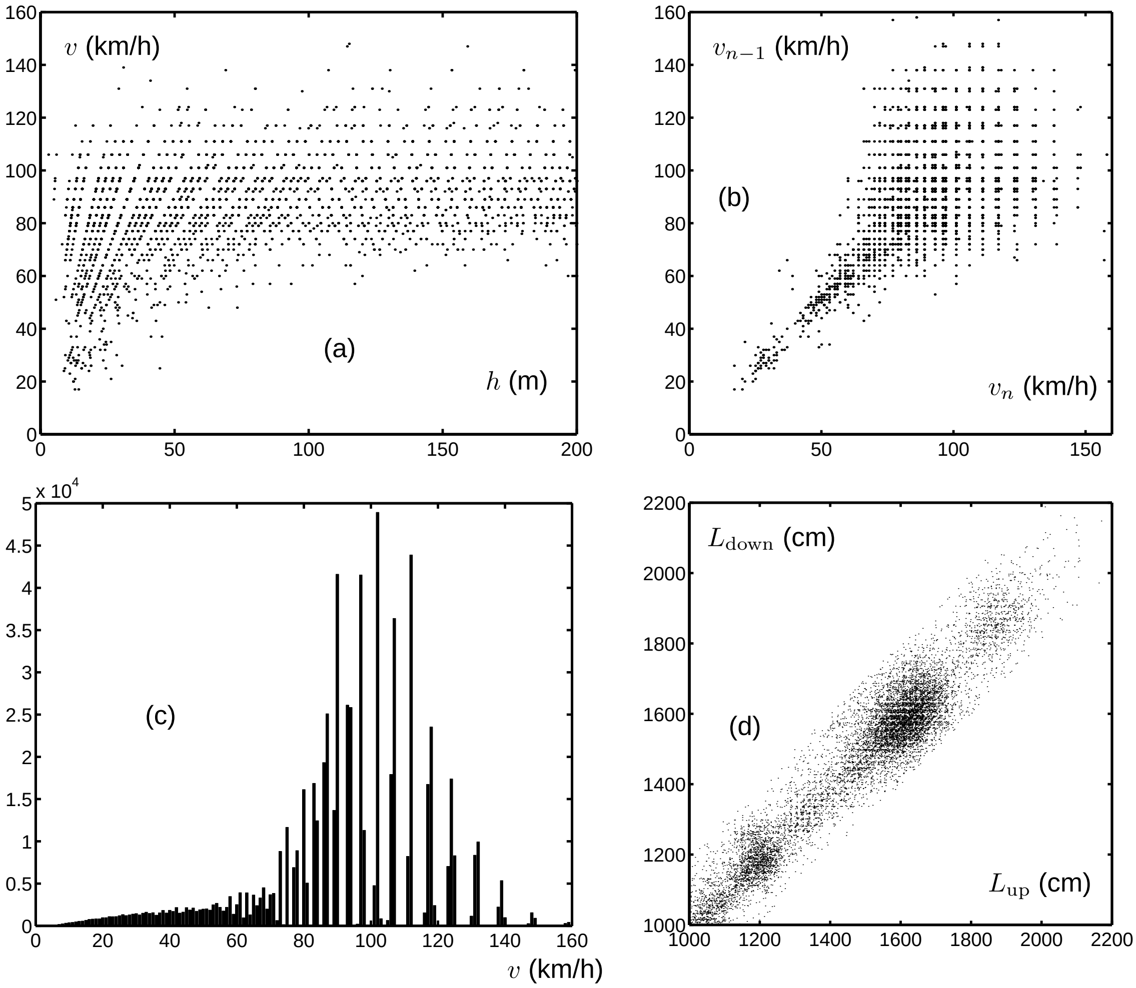

Figure 3: (a) Speed-headway data of individual vehicles. (b) Scatter plot showing the correlation of velocities of consecutive vehicles. The velocity of the leader is plotted against that of the follower. Note that the correlation is weaker for faster traffic.. We believe that an analysis of the skew about $v_{i-1} = v_i$ may lead to new statistical theories of platoons. The data from (a) and (b) are taken from lane 1. (c) Histogram of measured speeds at a single loop, with bucket size 1km/h. There is granularity and the measured speeds of fast vehicles are not accurate to the quoted 1km/h. (d) Scatter plot of measured length at downstream outstation against measured length at upstream outstation for a set of *definite match* lorries. The outstations are statistically identical, but inconsistent in their length measurements. This data might be used to obtain distributions of length measurement error.

all vehicles which pass the downstream outstation must have also passed the upstream outstation. Unfortunately, it does not follow that there is a complete 1-1 correspondence between the vehicle records of the upstream and downstream outstations: it is known that outstations sometimes miss vehicles, particularly in bumper-to-bumper traffic. It is thought that this failure rate is of the order of 1%.

The key to our analysis of length measurement errors is to remark that vehicles longer than, say, 10m, are rather infrequent. With such vehicles it is thus possible to establish a set of *definite matches* (concise definition to follow) where records at the upstream and downstream outstations must correspond to the same vehicle, because no vehicle of comparable length was present on that section of the motorway at that time. We call this process *vehicle re-identification*.

We can use a set of *definite matches* to produce a scatter plot of length measurements at the downstream outstation against length measurements at the upstream outstation: see Fig. 3(d). It is clear that the upstream and downstream length measurements for an individual lorry may occasionally disagree by up to two metres. However, since the statistical properties of upstream and downstream length measurements are almost identical, it is clear that the errors are due to a lack of consistency in the measurement of each individual outstation, rather than overall (e.g., physical) discrepancies between the outstations. Thus it may be possible to model the error in length measurement at an individual outstation by a random variable X which itself depends on the true length of the vehicle. The distribution of X might be inferred from Fig. 3(d), which contains information about (X_1, X_2) pairs, where we assume that X_1 and X_2 are independent and identically distributed.

Note that the techniques outlined above cannot tell us anything about the distribution of length measurement error for shorter vehicles, since these vehicles are so common that it is impossible to establish a set of *definite matches*. There is thus some scope in the future for controlled experiments on outstations using vehicles of known length, although it may be more appropriate to derive error distributions by analysing the microcontroller software.

The presence of such large errors in length measurement is extremely unfortunate. If length measurements were more accurate, it might be possible to re-identify *all* passing vehicles. This idea has been employed by Coifmann (2002) to calculate vehicles' travel times through a section of Californian freeway. Coifmann's method gives a maximum 60% re-identification rate. However, our main interest is in using re-identification to estimate lane changing rates, by comparing the order in which individual vehicles pass each outstation: this goal will require re-identification rates close to 100%.

The central idea of this research programme is that *vehicle re-identification* may proceed in a probabilistic manner, by seeking the most likely matches between upstream and downstream records. Fig. 4 shows short sequences of upstream and downstream length measurements taken from lane 1. The overall pattern of the data is very similar and it seems that we can match most of the sequences with some confidence.

We need to define some terminology before we proceed further. Firstly, we say that we have a *possible downstream match* of an upstream record, if

- the measured downstream length is within some tolerance of the measured upstream length; and
- the downstream timestamp is within some tolerance of that predicted by considering the upstream timestamp, the upstream measured speed, and the distance between the outstations.

A *possible upstream match* of a downstream record is defined similarly, except that one considers a backward forecast of the upstream timestamp using the downstream timestamp and downstream measured speed. We say that a pair of records are a *possible match* if the downstream record is a *possible downstream match* for the upstream record, and the upstream record is a *possible upstream match* for the downstream record. Further we say that two records are a *definite match* if the downstream record is the *unique possible downstream match* of the upstream record, and the upstream record is the *unique possible upstream match* of the downstream record.

It would clearly be desirable if we could re-identify 100% of records as *definite matches*. However, due to the length measurement error we described above, this is not possible.

The principal question concerns the size of the length and timestamp tolerances mentioned above. If these tolerances are too tight, then it is possible that many records will have no possi-

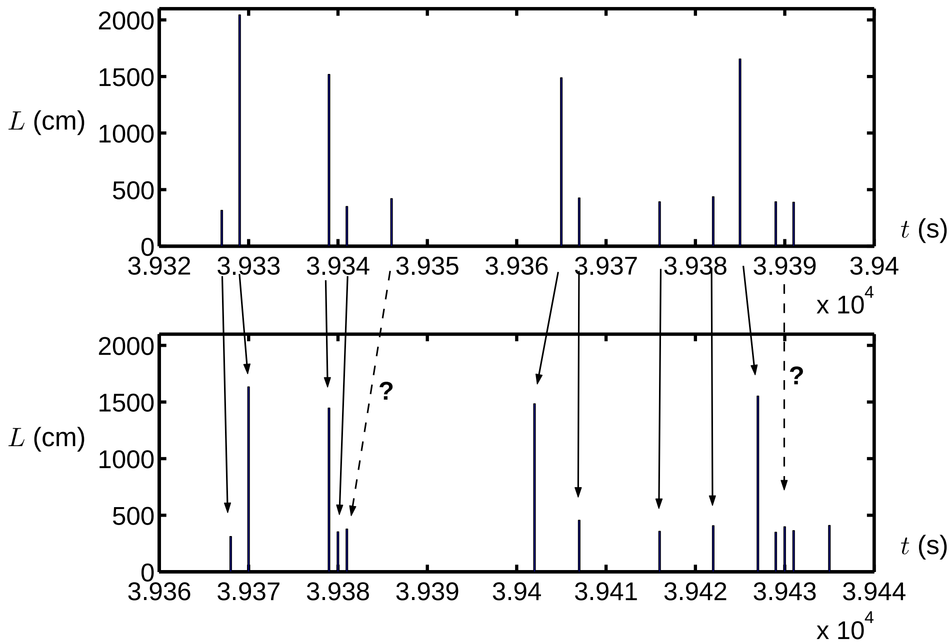

Figure 4: Short sequences of vehicle length measurements against time, taken from lane 1 at (top) the upstream outstation and (bottom) the downstream outstation. Note that there is an offset of 40s between the two graphs. It appears possible to re-identify most vehicles, despite the length measurement errors.

ble matches at all. In fact, as mentioned earlier, the outstations fail to record some vehicles, so that even under loose tolerances, some records will have no possible matches.

However our aim is to re-identify as many vehicles as possible, even if this process includes a probabilistic element. Thus our strategy is to set loose tolerances where typical records will have more than one possible match. The idea is then to score each possible match, and find a set of pairings which overall has the best score. However, our basic data set has over 600,000 records at each outstation, and we cannot expect to match such long sequences in one step.

Our approach therefore is to try and break sequences of traffic into smaller subsequences and to seek optimal matches between these. The overall strategy is to use long vehicles (usually defined as over 10m) as reference points which assist in breaking up sequences. A first step therefore is to obtain *likely matches* (somewhat weaker than *definite matches*) for large numbers of long vehicles throughout the data set.

The mathematical formulation of the full process is extremely complicated and is still under development. Here we illustrate the idea with an example of the search for likely long vehicle matches, which has most of the key features of the more general search for shorter vehicles once long vehicle reference points have been found.

The first step is to choose a long vehicle record in the upstream data set. We then find all possible downstream matches, and for each of those possible downstream matches, we find all possible upstream matches; for each of those possible upstream matches we find all possible downstream matches, and so on. If we only allow matches with records whose measured length is above a certain threshold, and if we set error tolerances sufficiently tight, the process described will terminate: i.e., we will end up with sets D and U of downstream and upstream records respectively, such that all possible downstream matches of all elements of U are contained in D, and similarly all possible upstream matches of all elements of D are contained in U. For a real example, see the data of Fig. 5, where the process terminated with the $|D| = |U| = 4$.

The next step is to calculate possible (bi-directional) matches between D and U; i.e., to identify all pairs of elements $d \in D$ and $u \in U$ such that u is an upstream match of d, and d is a downstream match of u. This gives a possible match matrix, which for the Fig 5 example takes the form

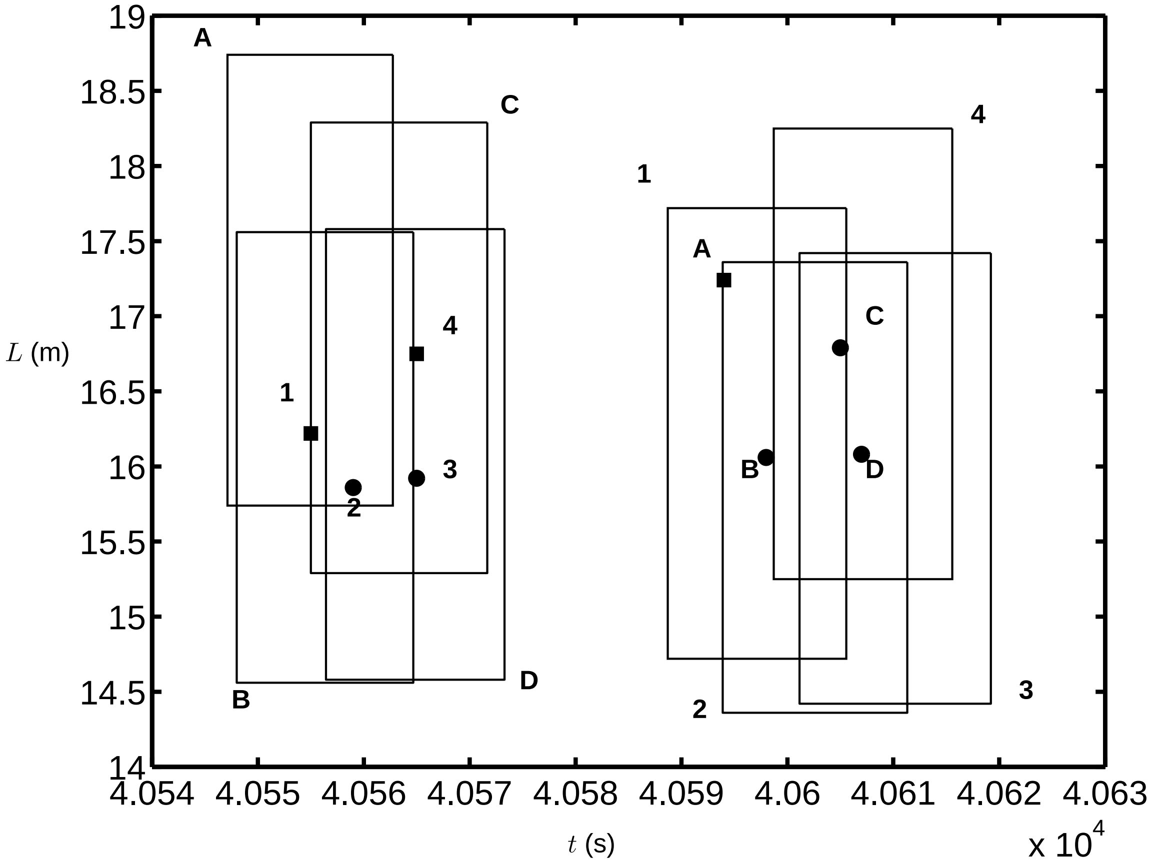

Figure 5: This figure plots measured lengths against timestamp and illustrates how the principle of possible matches may be used. The length error tolerance is ± 1.5m and the timestamp tolerance corresponds to a $\pm 20\%$ deviation in average speed between the loops, compared to that which is measured. The leftmost group of circles (lane 1) and squares (lane 2) labelled **1**, **2**, **3**, **4** are upstream records and the group of rectangles labelled **1**, **2** etc. towards the right of the figure represent the tolerances which downstream records (the circles and squares labelled **A**, **B**, **C**, **D**) must fall within to be considered *possible downstream matches*. Likewise, the leftmost group of rectangles **A**, **B**, **C**, **D** represent the tolerances which upstream records must fall within to be considered *possible upstream matches* for the respective downstream records. Therefore the possible downstream matches of **1** are **A**, **B**, **C**; of **2** are **A**, **B**, **C**, **D**; of **3** are **C**, **D**; of **4** are **C**, **D**. Likewise the possible upstream matches of **A** are **1**, **2**; of **B** are **1**, **2**; of **C** are **2**, **3**, **4**; of **D** are **2**, **3**, **4**. It follows that **A** and **B** are each possible matches of **1** and **2**; further **C** and **D** are each possible matches of **2**, **3** and **4**. If we assume that each record has at least one possible match, then matches of **C** or **D** with **2** are impossible, because we seek a bijection between $U := \{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}\}$ and $D := \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ It follows that we must calculate scores to decide which way round **A** and **B** match **1** and **2** and which way round **C** and **D** match **3** and **4**. Different scoring formulations are possible, but if one considers the distances of points relative to the centres of the required tolerance rectangles, then **A** matches **1**, **B** matches **2**, **C** matches **4** and **D** matches **3**. It is clear that the vehicle corresponding to record **4** has been undertaken by that corresponding to **3** and **4** has subsequently pulled in to lane 1.

	A	**B**	**C**	**D**
1	1	1	0	0
2	1	1	1	1
3	0	0	1	1
4	0	0	1	1

Then we need consider which amongst the possible matches are the most likely. To do this we choose and calculate a scoring function. For example, a match might be scored according to the appropriately weighed sum of (i) the discrepancy in length measurements of the two records, (ii) the difference of downstream timestamp and the forecast obtained from the upstream record, and (iii) the difference of upstream timestamp and the backward forecast obtained from the downstream record.

In the simplest case where D and U have the same number of elements, we seek the pigeon-holing map Π (i.e., the bijection) between the two sets which minimises some norm (e.g., the sum) of the scores of each of its matches. As illustrated in the example of Fig. 5, the search for a bijection may immediately eliminate certain combinations of possible matches and massively reduce the search space.

The search space of bijections consists of permutation matrices which are invariant when ANDed with the possible match matrix given above. Under most natural scoring functions, the Fig. 5 example yields the optimal bijection

$$\Pi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{2}$$

In the caption of Fig. 5, we describe how lane changing information is inferred from this result.

3 Aerial video analysis

In this section we announce work on the automatic image processing and analysis of video pictures of highway traffic which have been taken from a helicopter flying above the motorway. These pictures were originally filmed by TRL as part of a Highways Agency funded project to study driver behaviour at motorway merges. However, we believe that modern data analysis tools will be able to extract much more information.

In particular, since the field of view is several hundred metres (see Fig. 6), it should be possible to track vehicles for several seconds and obtain displacement time series with a resolution of 25Hz. This time series can then be differenced to yield new time series for velocity and acceleration. This is precisely the information required to fit the details of microsopic car-following laws which has not previously been available.

In fact, Low and Addison (1999) have pioneered the use of aerial pictures to fit car-following models. As part of an EPSRC funded project, they chartered a helicopter and followed vehicle pairs down the motorway. They subsequently measured vehicles' displacements by hand and attempted to fit some simple nonlinear dynamics to the results. The goals of our work are much more ambitious:

- Firstly, we want to analyse large numbers of interacting vehicles, so that we may understand the full range of driver variablity and its effect on macroscopic dynamics. It follows that we would rather not measure vehicles' image coordinates by hand, and we would like as many steps of the process as possible to be carried out automatically in software.

- Secondly, we would like to use existing aerial pictures which were originally recorded for other purposes than our own. These pictures may be less than ideal: e.g., the cars may appear small and rather too far away, or the camera angle may be bad, and high winds may also buffet the helicopter and lead to pictures which shake.

Our approach is based on standard modern image processing techniques; see Gonzalez and Woods (1992) for an introduction. The process is as follows:

Figure 6: Screen capture from our automatic image analysis software package. The operator has selected a vehicle, top picture approximate coordinates (490,150) and the software has automatically tracked the vehicle for 3 seconds (bottom picture).

- We copy the pictures on to hard disk using a video capture card. This gives 25 bitmap files per second. Each bitmap file has 24 bit colour and about 600 rows and about 700 columns.
- We throw away all the even numbered pixel rows. This is to eliminate the undesired *combing* effect which arises because the video camera is interlaced.
- The operator looks for prominent (stationary) features which are present in as many video frames as possible. The operator selects rectangular regions around each of these *reference points* in the first image frame of the sequence, and the software remembers the bitmaps inside these rectangles.
- The software package automatically tracks the pixel coordinates of each of the reference points in every video frame. (See Fig. 7 for an illustration of this process.) This tracking is achieved by seeking a least squares best fit position of the initial stored bitmap to each new video frame. The initial position for the search is usally the pixel coordinates of the reference point in the previous frame. For computational efficiency, a maximum size (number of pixels) for the search window must also be defined, and if the reference point moves more than that number of pixels between frames, the software will usually fail to track the reference point.
- The pixel coordinates of all (stationary) reference points together will allow one to define for each frame a transformation from pixel coordinates to a displacement defined in terms of the relative (real) displacement of the reference points. (All reference points and vehicles must be approximately co-planar in real space.) When tracking moving vehicles, this allows one to eliminate the component of apparent motion due to the motion of the camera. Absolute distances can be fixed by measuring the length in pixels of some object in the frame whose size is known, e.g., the lane markings.
- The least squares matching process described above is used to follow vehicles through the sequence of video frames: see Fig. 6. Knowledge of the reference points' pixel coordinates

and vehicles' velocities enables one to obtain refined guesses for the initial search position. After pixel coordinates of vehicles have been obtained, real displacement coordinates may be obtained using the transformations described above.

The last part of this work is at the limits of what automatic image analysis can presently achieve, and work is still in progress. However, at the UTSG conference we hope to announce the first large data set produced by automatic image analysis.

4 Conclusion

In this paper we have reported progress on our analysis of two new microscopic highway traffic data sets, namely (i) MIDAS IVD and (ii) aerial video pictures. In our view, the analysis of microscopic data is essential if highway simulation packages are to be improved.

We would like to thank the Highways Agency for allowing us to use these data sets (which were originally collected for other purposes) free of charge.

In the future, we feel that there is scope for the deliberate collection of very much more comprehensive data than that described here. In particular, use of triple inductance loops, or clever double loop algorithms, could lead to point measurements for acceleration. We believe that there is also scope for reducing vehicle length measurement error by improving the microcontroller software. Our belief is that comprehensive data collection, in combination with modern data analysis techniques, will in time lead to a thorough quantitative understanding of all aspects of driver behaviour.

References

Abou-Rahme N. and White J. (1999). *Replicating shockwaves within a classical car-following model*, in: TRL annual research review 1999, pp73–82.

Bando M., Hasebe K., Nakayama A., Shibata A. and Sugiyama Y. (1995a). *Dynamical model of traffic congestion and numerical simulation*, Phys. Rev. E **51** pp1035–1042.

Bando M., Hasebe K., Nakanishi K., Nakayama A., Shibata A. and Sugiyama Y. (1995b). *Phenomenological Study of Dynamical Model of Traffic Flow*, J. Phys I France **5** pp1389–1399.

Coifman B. (2002). *Estimating travel times and vehicle trajectories on freeways using dual loop detectors*, Transpn. Res.-A **36** pp351-364.

Dixon C., Rees T., Sutch T., Lunt G. and Quick J. (2002). *Monitoring of the M25 controlled motorway 2000–2002*, TRL unpublished report, no. PR/T/077/02 T020

Gipps P.G. (1981). *A behavioural car-following model for computer simulation*, Transpn. Res.-B **15** pp105–111.

Gonzalez R.C. and Woods R.E. (1992). *Digital Image Processing*, Addison-Wesley.

Low D.J. and Addison P.S. (1999). *Traffic dynamics investigated using mathematical modelling and aerial video filming*, Proc. 32nd ISATA conference, Vienna, 1999.

Wilson R.E. (2001). *An analysis of Gipps's car-following model of highway traffic*, IMA J. Appl. Math. **66** pp509–537.

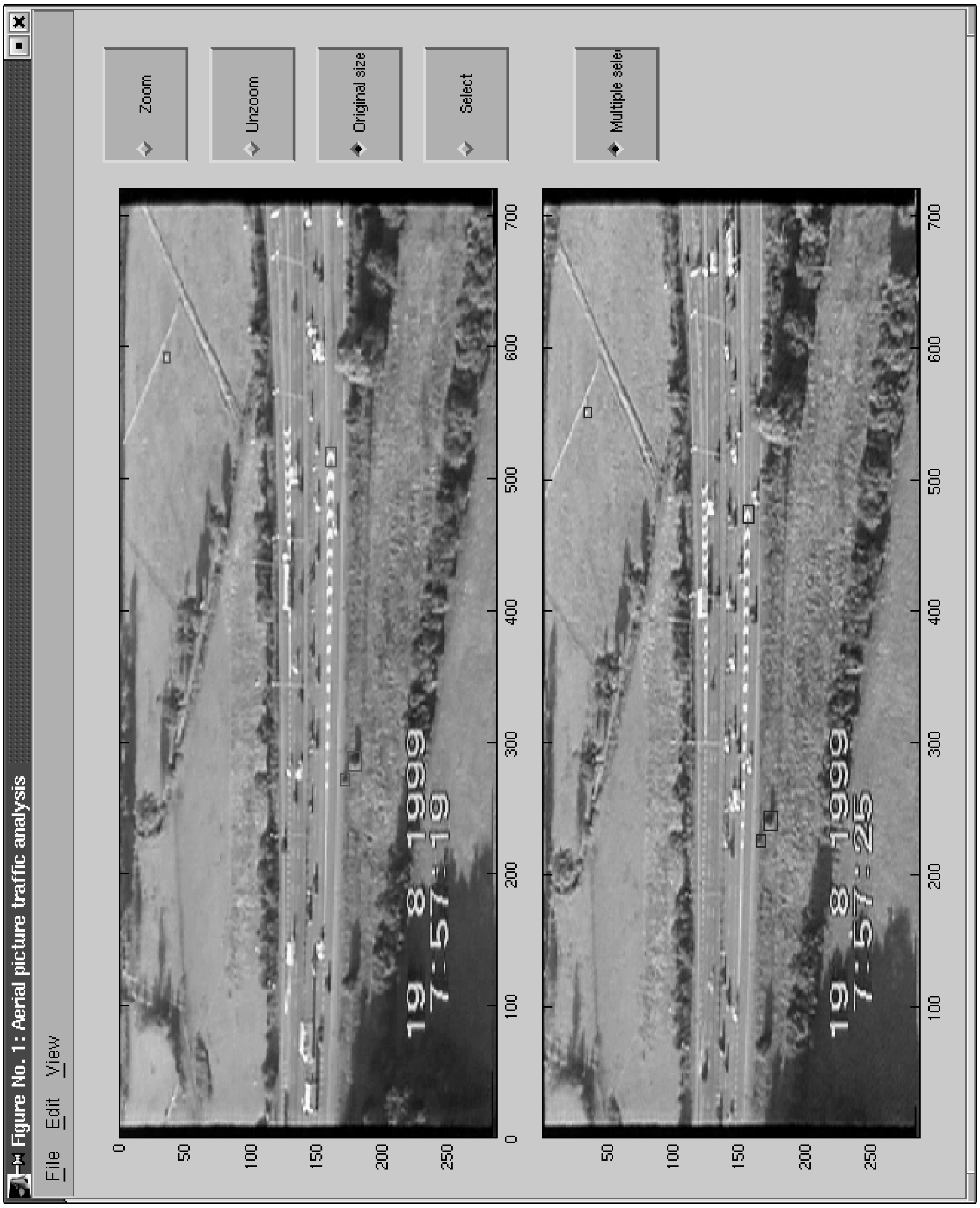

Figure 7: Screen capture from our automatic image analysis software package. (View sideways.) The operator has selected small rectangular regions of the top picture which contain easily identifiable features (small bushes, chevrons etc.). These features are stationary and are to be used as *reference points*. Our software has automatically tracked the features through 5 seconds of video: see bottom picture. Although the features are stationary, their pixel coordinates change due to the motion of the helicopter. Tracking stationary reference points allows one subsequently to factor out that component of vehicles' apparent motion which is due to the motion of the helicopter.