UNIVERSITY OF SOUTHAMPTON

A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

A PROBABILISTIC MODEL FOR THE ORIGIN OF MOTORWAY SHOCKWAVES

BY

NABIL F. ABOU-RAHME

TRANSPORTATION RESEARCH GROUP
DEPARTMENT OF CIVIL & ENVIRONMENTAL ENGINEERING
2003

to

my parents
with gratitude

and

my grandparents
whose lives were an
inspiration

CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

UNIVERSITY OF SOUTHAMPTON
FACULTY OF ENGINEERING AND APPLIED SCIENCE
DEPARTMENT OF CIVIL ENGINEERING

ABSTRACT

DOCTOR OF PHILOSOPHY

A PROBABILISTIC MODEL FOR THE ORIGIN OF MOTORWAY SHOCKWAVES

by
Nabil Farid Abou-Rahme

From stop-start driving through to stationary queues, traffic congestion in its various forms is becoming an increasingly common feature on the motorway network. Flow-breakdown on motorways is a complex phenomenon that occurs under a combination of specific traffic conditions and perturbations in the flow.

The context of this thesis is the value of being able to pre-empt the upstream propagation of the sequential braking pattern (known as a shockwave) through dense traffic, thereby providing an opportunity to influence the traffic using speed control. The current body of literature describes flow-breakdown in various forms, including car-following behaviour, merge conflicts, gap acceptance, and fluid models for shockwave propagation. However, the area of shockwave origins (seed-points) is relatively unexamined.

A fresh representation of shockwaves was obtained from traffic data on the M25 London Orbital motorway, and this led to a deeper investigation into the properties and mechanisms behind the formation of shockwaves. This investigation comprised a six-month survey of traffic data (accompanied by video footage) in order to examine the variety and distribution of events that initiate shockwaves.

The implementation of automatic variable speed-limits on the motorway network raises the possibility for dynamic shockwave damping. A simple feasibility model is presented in this thesis, showing the importance of being able to predict the start of a shockwave in order to provide adequate time to influence drivers upstream of the origin point.

Research then focussed on how to model the origin of the shockwave, with a primary emphasis on statistical inference. A probabilistic model for the origin of motorway shockwaves is presented, requiring a detailed description of conditions present at the commencement of a shockwave. Detection of seed-points was therefore critical to the implementation of this probabilistic model.

Examination of existing methods of incident detection within the data set showed that whilst they detected the body of a shockwave effectively they did not extend to finding the origin point. Software has therefore been developed to identify seed-points, and descriptions of the proposed algorithms used in the software are presented and assessed on their ability to identify the origin of shockwaves. This study has shown that it is possible to identify the location of seed-points by classification using traffic characteristics. The primary variables used were average speed and occupancy (a proxy for headway) as obtained from MIDAS detectors on the M25 Controlled Motorway.

Finally, as a proof of concept, the probabilistic method was applied to the detected data sets in order to demonstrate the calculations involved. Probability maps were obtained for individual months and the implications were reviewed. Accurate calibration of this model is outside the scope of this present study, and points to a further research opportunity.

The true value of this research will be in its application to intelligent transport systems which combine control algorithms with vehicle-based speed control systems.

ACKNOWLEDGEMENTS

I was first introduced to the subject of shockwaves by **Professor Benjamin Heydecker** whilst studying for my Masters Degree in 1994. His clear presentation of the knowledge to date drew my interest and I became fascinated by the relatively untouched area of how and why shockwaves occur in the first place.

On commencing employment with **TRL Limited**, I began working with volumes of traffic data from the M25 Controlled Motorway. The shockwave patterns were evident in the analysis and once again the elusive origin (seed) points caught my attention. Since I had been recruited as part of a PhD Programme, it seemed an ideal opportunity to follow through on some of the questions in my mind.

I would like to gratefully acknowledge **Professor Mike McDonald** at the University of Southampton for his supervision and guidance throughout this study. I also acknowledge **Dr Phil Bly** at TRL Limited for the opportunity provided through the PhD Programme, and in particular thank **Peter Still** and **Dr Phil Hunt** for taking an active interest in my research.

Thank you to **Dr Jason Cheung** for advice on software issues and coding of incident detection algorithms, to **Dr Joanna White** and **Dr Babul Baruya** for guidance on the application of statistics in this study, and to **Dr Ahmer Wadee** at Imperial College London for his wisdom, friendship... and regular enquiries about my progress!

I also tip my hat to **Brian Harbord, Stuart Beale** and **Robert Stewart** at the Highways Agency for taking an interest and supporting various aspects of my research within their projects. I look forward to discussing further research!

Apart from being able to make a contribution and the satisfaction that comes through completed work, perhaps the most enjoyable aspect of my research was the opportunity to speak about shockwaves on the BBC "Gridlock" programme in August 2000. Presenter **Nick Ross** made two astute observations – first that shockwaves "drive you mad" and later that "Nabil is fascinated by traffic jams". After *almost* ten years of fascination I must protest that these two statements are not necessarily connected!

Finally it has not been easy completing this research whilst coping with the demands of full-time employment. The equation of life has many variables to be optimised, so my gratitude goes above all to family and friends for your support, patience and understanding.

"The end of a thing is better than its beginning... much study is wearisome to the flesh, and of the making of many books there is no end".

– Ecclesiastes 12

1 AN INTRODUCTION

1.1 FLOW-BREAKDOWN

From stop-start driving through to stationary queues, traffic congestion in its various forms is an increasingly common feature on the motorway network. Every day, thousands of commuters use the motorway and trunk-road network to carry out journeys to and from their place of work or study. Demand for motorway travel is rising as fast as 3% per annum on motorways such as the M25, but the emphasis of current government policy is on providing better traffic control to existing roads rather than building new ones.

The breakdown of traffic flow is a complex phenomenon, occurring under a combination of specific traffic conditions and perturbations. When the source of the congestion appears to concentrate around a particular location the cause can often be traced to the effect of local geometry on traffic behaviour, which can provide a level of disturbance sufficient to trigger flow-breakdown. Congestion is not always initiated by an immediately obvious cause such as an accident or roadworks, in some cases traffic moves (with little warning) from a volatile high-speed high-density state into a flow-breakdown state. Once the speeds do begin to drop, it can take some time before traffic returns to a freely flowing state again.

Flow-breakdown has distinct economic implications in terms of lost time and missed opportunities. Unlike a train journey, time in the car cannot be easily transferred to other tasks, and late deliveries in the logistics industry can be quite costly. Congestion brings increased emissions through erratic driving patterns and an increase in the level of driver stress and frustration (due to the lack of control an individual has over the length and progress of their journey).

One study on motorway delays showed that for a particular stretch of busy motorway during an average weekday the recurrent congestion was responsible for around 3000 vehicle-hours delay (Still, Abou-Rahme and Rees, 1997). This indicator is explored further in Chapter 3, but the value signifies an aggregation of delay experienced by those drivers on the

motorway unable to travel at their desired speed. By comparison, the vehicle-hours delay on non-typical days, (such as when a serious accident has occurred), can reach values of 9,000 or more. The same study observed that over a period of several months the delays due to recurrent congestion were about three times larger than the congestion directly attributed to incidents.

The problem of flow-breakdown is influenced by a number of significant factors such as distribution of speed, and the density and volume of traffic. Flow-breakdown appears to be the result of a combination of fixed geometric features (such as curvature, gradient, or junction layout) and the variable characteristics of traffic passing through. Composition by vehicle type can alter the degree of severity for flow-breakdown, whilst the distribution of spacing between vehicles (headway) can affect the overall stability of traffic leading up to flow-breakdown.

One way of visualising flow-breakdown along a stretch of motorway is to think in terms of a hurdle-race. Runners come across the first hurdle and either pass unhindered, stumble, or fall over. If runners pass a hurdle successfully they then continue forward to tackle the next hurdle. The analogy here is that each hurdle is a potential pitfall only until it is cleared. In similar fashion clusters of vehicles traverse a section of motorway, passing through locations that are conducive to flow-breakdown. If conditions at one such location are favourable then traffic proceeds unhindered to negotiate the next potential location.

Flow-breakdown, especially during peak hours, is not always initiated by an immediately obvious cause. In any given unstable flow there exists a phenomenon known as a shockwave (described further in Chapter 2), a change of state that propagates upstream. Those vehicles involved in the initiation or early stages of a shockwave are often unaware of its existence.

Figure 1 illustrates the causality behind the occurrence of a shockwave, based on the logic described above.

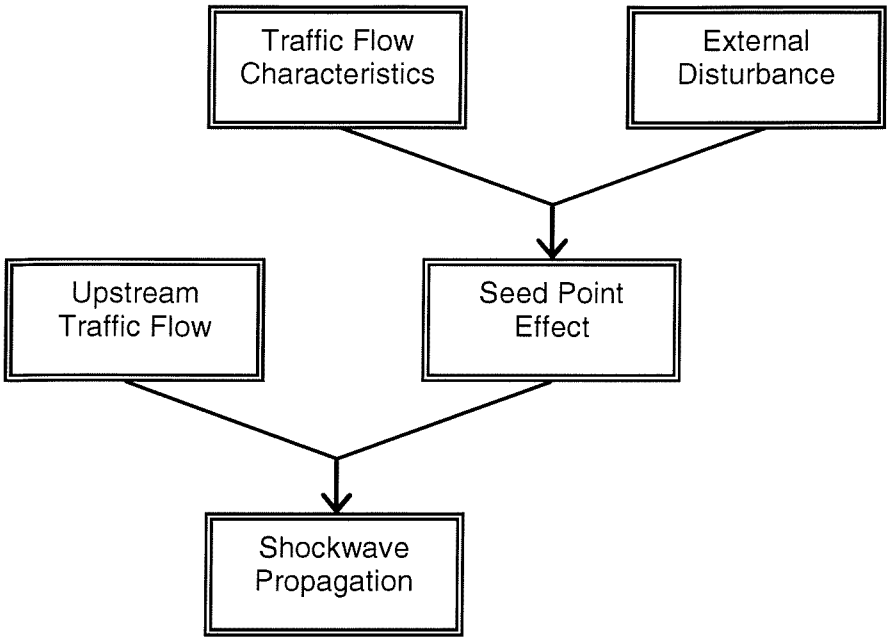


Figure 1 – Cause and effect representation for shockwave propagation

A temporary restriction to traffic flow may have benefits to vehicles downstream of that bottleneck. An example of this which can be readily observed is the practice of slow overtaking by heavy vehicles. The overtaking vehicle moves up with relative ease alongside the vehicle it is passing and then matches its speed for a short interval. This creates a temporary bottleneck (particularly effective on a dual carriageway), thus giving congestion downstream a little more time to recover without being fed by a relentless flow of traffic. Observation of traffic patterns suggests that drivers recovering from a shockwave are dispersed, but eventually regroup with a headway distribution which makes them vulnerable for further shockwave propagation (see Dixon, Harbord, and Abou-Rahme, 2002).

In order to make a difference to traffic behaviour, a control system would need to make some prediction of flow-breakdown rather than just detect it after the event. A number of successful approaches to incident detection are already available (some of which have been used as part of the present study), but this can be taken a step further by investigating early warning signs in the traffic data.

1.2 STUDY OBJECTIVES

The primary aim of this research has been to better understand the origin of shockwaves in motorway traffic and hence, potential remedial actions.

More specifically, the objectives of this research were to:

- review the body of literature related to this subject;
- identify the occurrence of shockwaves in a particular archive of real traffic data, and locate their seed-points;
- examine the traffic conditions prior to the occurrence of flow-breakdown;
- develop and evaluate a conditional probabilistic model in order to describe the data and provide a posterior distribution for the occurrence flow-breakdown;
- propose future applications of the model derived.

1.3 METHOD OF APPROACH

The Road Traffic Act of 1991 introduced the legislative framework necessary for variable mandatory speed-limits on motorways and the use of photographic evidence for speed enforcement.

Using the Motorway Incident Detection and Automatic Signalling (MIDAS) system deployed on the M25 as the basic infrastructure for the new scheme, the Highways Agency developed an automatic speed-control environment known as 'Controlled Motorways'. A key feature that differentiates this road environment from a conventional motorway is the use of mandatory speed-limit signals (see Figure 2) rather than the advisory speed-limits displayed on traditional motorway signals.

The Controlled Motorway Pilot Scheme covers the south-western quadrant of the M25, extending from Junction 10 (A3) to Junction 15 (M4), a total of around 22km. Traffic information is collected by means of inductive loops and processed by various algorithms at the local subsystems. Congestion alerts are then processed in order to determine the most appropriate speed-limit to be displayed.

Figure 2 – Variable Speed Limits in operation as part of Controlled Motorways on the M25 London Orbital between J10 and J16

The driver-interface consists of Variable Speed-Limit displays mounted on gantries at regular intervals throughout the section. Each gantry has a number of speed cameras mounted on the reverse side in order to enforce the speed-limit displayed.

This stretch of the M25 is one of the busiest motorways in Europe, with flows now exceeding 200,000 vehicles per day. Traffic data (averaged over one-minute intervals) has been available from MIDAS detectors on a daily basis since 1995. Data quality was sometimes affected by faulty loops or outstation problems, so basic filtering was needed prior to analysis. Within the data are hundreds of occasions of flow-breakdown, and information about the corresponding upstream and downstream conditions. Such information has been used in this thesis to provide valuable insight into the traffic behaviour observed.

The M25 Controlled Motorway section has therefore been chosen as the key study area for this thesis.

1.4 STRUCTURE OF REPORT

This chapter has given a brief introduction to the breakdown of traffic flow on motorways and described the main study site for the thesis, namely the M25 Controlled Motorway.

In Chapter 2 the shockwave phenomenon is described in more detail looking particularly at its classical definition and characteristics in terms of origin and propagation. The chapter concludes with a survey of the body of literature that relates to this field, and a logical progression as to why this present approach was identified as suitable for research.

In Chapter 3 an examination of the feasibility of dynamic shockwave damping sets the context and relevance of the problem being studied in this thesis. A six-month study into the causes of congestion on the M25 was conducted, and the results are presented to demonstrate understanding of the nature of seed-points. Chapter 4 sets out a number of alternative approaches to modelling the seed-points based on a qualitative understanding of traffic behaviour and the intrinsic complexity of microscopic models. The description includes reference to car-following behaviour and how, when combined with appropriate lane-changing models, it can lead to seed-point creation. The probabilistic model is identified as the most appropriate for application. Input required for each component of the model is explained, and the potential to harness the predictive power of traffic forecasters is discussed.

Having highlighted the importance of detecting the seed-points in order to have some input for the probabilistic model, the thesis then concentrates in Chapter 5 on the methods of detection. The chapter demonstrates how the seed-points were identified and analysed in the traffic data, summarising the overall methodology. Software has been developed to identify seed-points, and descriptions of the proposed algorithms used in the software are presented and assessed on their ability to identify the origin of shockwaves. The relationship between speed and vehicle headway is explored using real data of a higher resolution than the standard detector outputs. The added

insight provided a useful guide as to how the seed detection software should be applied.

Results of the seed-point detection experiment are presented in Chapter 6, along with some simple analysis of the traffic data. The probabilistic method was applied to the detected data sets in order to demonstrate the calculations involved. Accurate calibration of this model is outside the scope of this present study, and points to a further research opportunity. The output is given in the form of a posterior distribution describing the probability of a shockwave occurring for various combinations of speed and traffic density.

Chapter 7 presents the conclusions of the research and looks at the way forward, describing possible applications of the work and suggestions for its development in the context of a rapidly developing industry. The application of this work would ultimately be in the ability to respond appropriately to conditions which are highly likely to produce flow-breakdown, and to control flow upstream of a shockwave to ensure the shockwave itself will decay and not propagate. The work also finds application in the enhancement of traffic forecasting models where the effects of seed locations need to be taken into account.

References and appendices are included at the end of the document. Appendix A gives further background and results from the monitoring of Controlled Motorways on M25 J10–15. Appendix B discusses methods of data collection alternative to loop detectors. Appendix C summarises the car-following equations discussed in Chapter 4. Appendix D explores the issues of accuracy for data from loop detectors and the differences between point and section-based data collection. Appendix E gives further results in support of the arguments put forward in Chapter 6. Appendix F gives relevant software listings and Appendix G contains a glossary for terms used in this thesis. Appendix H expands on an alternative approach to deriving the speed of shockwave propagation, giving supporting material to Chapter 2.

2 SHOCKWAVES IN TRAFFIC

2.1 INTRODUCTION TO SHOCKWAVES

This chapter introduces the concept of shockwaves and their contribution to the phenomenon of flow-breakdown. A qualitative description of the shockwave as a moving boundary between two states of traffic flow is reviewed.

Thls is followed by a comprehensive review of the current body of literature, from which a number of papers are identified forming the foundation for the hypothesis set forth in this study.

2.2 DESCRIBING A SHOCKWAVE

During congested traffic flow, the traditional assumption of independent and random arrivals is no longer valid. Under such conditions traffic is often affected by conditions further ahead in the direction of travel (downstream), and the achievable speed becomes restricted. In general terms, a shockwave is a wave which moves through a fluid faster than a change would normally be propagated. In the context of traffic flow a shockwave can be described as a high-density zone which moves against the direction of flow (upstream) over a period of time. The boundary or discontinuity between this region of high-density and other regions of lower density is also commonly termed 'the shockwave' as the primary effects occur there. Such a discontinuity can sometimes occur well before the overall demand on a link has reached or exceeded the physical capacity.

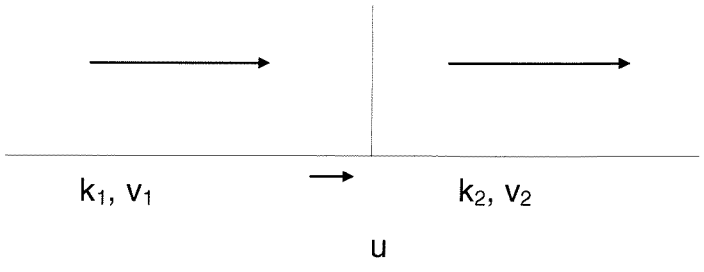
The origin point from which a shockwave propagates can be described as the 'seed-point'.

Shockwaves are broadly characterised by a constant speed of propagation as well as by an increasingly large and abrupt drop in traffic speed as they pass a particular location. The severity of a shockwave, defined in terms of the degree of reduction in speed, often increases with distance upstream of the seed-point. Even minor overreaction from consecutive drivers will bring about an amplification of the braking effect over time.

If propagation is sustained then there comes a point at which the shockwave is no longer just causing a fluctuation in traffic speeds but is also bringing vehicles to a complete standstill. Once this level of severity is reached, the shockwave begins to grow in another dimension, namely the duration for which this effect remains at a particular location. Cumulative delay is introduced by a range of readily observable behaviour by drivers, for example the need to get into gear and pull away again, or perhaps waiting until traffic ahead has moved more than just a couple of meters. Each delay is small and sensible but the accumulation defines the time for which a shockwave acts at a particular location. We have therefore identified five characteristics of a shockwave which can assist our understanding:

- physical location of the shockwave origin point (a seed-point)
- direction of travel for the boundary between traffic states
- speed of propagation (assumed constant but varies slightly)
- severity of impact (speed reduction across the boundary)
- duration of impact (time until vehicles achieve speeds similar to those before encountering the shockwave).

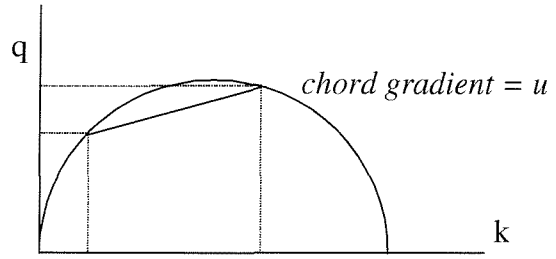
A simple scatter graph of traffic data for density (k), velocity (v) and flow (q) shows them to be related in the approximate form $q = kv$. On a macroscopic scale, a platoon of traffic (with properties k_1, v_1) approaches another platoon (with properties k_2, v_2) such that $k_1 < k_2$ and $v_1 > v_2$. We can represent this as follows:



The abrupt increase in density in the direction of travel between the two regions on the carriageway defines the shockwave. This boundary travels at a velocity (u) that can be defined in terms of the properties of the associated vehicles, given that conservation of flow holds true across the boundary.

Flow in: $q_i = k_1 (v_1 - u)$

Flow out: $q_o = k_2 (v_2 - u)$



Solving these equations for boundary velocity (u) gives:

$$u = \frac{(k_1 v_1 - k_2 v_2)}{(k_1 - k_2)}$$

assuming that $q = kv$ holds true for the specific traffic data being observed. Although flow at the boundary is conserved, the speed of propagation is the ratio of the differences in flow and density between the two platoons. Since $k_1 < k_2$ and $v_1 > v_2$ the value of u is often negative suggesting backward propagation against the direction of travel. Although other combinations are possible, this study does not examine them.

Figure 3 uses vehicle trajectories to present an example of shockwave propagation. The shockwave becomes apparent when consecutive vehicle journeys are examined following a minor disturbance to traffic with relatively high flow and density. In Figure 3 the shockwave can be seen as the shadow region or ridge, travelling from left to right upstream. Each vehicle has a similar response to the braking pattern of the vehicle in front, this 'chain reaction' passes back upstream the message to brake. Once a vehicle has responded to the shockwave, it then appears to return to its original speed.

The width of the diagonally propagating ridge can be seen to widen over time, suggesting that the degree of braking is amplified by the successive reaction of drivers as the wave propagates. Once a shockwave reaches the end of the platoon, the message becomes ineffective, since there are no more cars behind to continue to react to the braking pattern. The shockwave has dissipated.

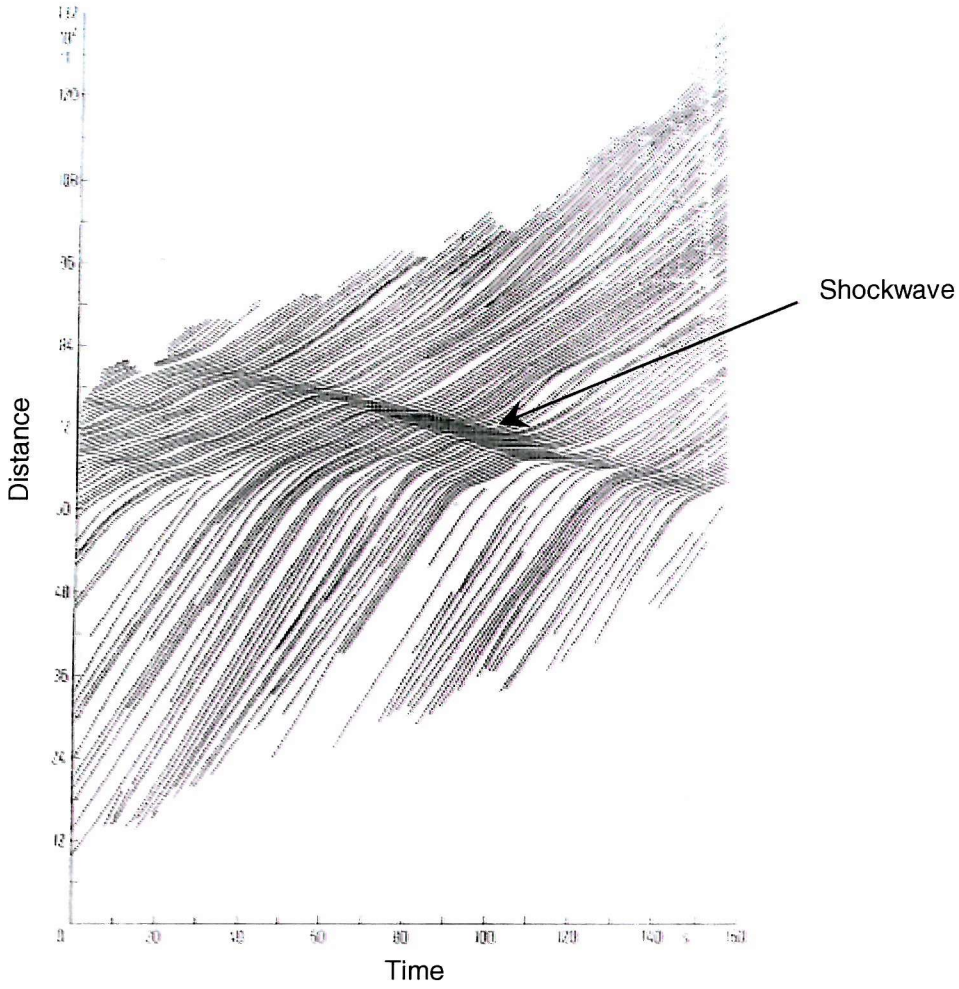


Figure 3 – A shockwave defined by consecutive vehicle trajectories, after Leutzbach (1988)

The Motorway Traffic Viewer (MTV) facility (developed at TRL Limited within a project led by the author) provides a variety of graphical representations for loop and signal data. As part of the M25 Controlled Motorway monitoring it was possible to identify shockwaves within the traffic data, and demonstrate that the speed of propagation is approximately 19km/h (with variation from 18–23km/h). If this boundary velocity (u) can be assumed constant for the propagation of the shockwave and the relationship $q = kv$ holds true to the data set then the difference in flow across the boundary is directly proportional to the difference in density. A fuller discussion of how the boundary velocity equation can be applied to the available MIDAS data is given in Appendix H.

Often, those vehicles involved in the initiation or early stages of a shockwave are unaware of its existence, the direct effect of a shockwave being felt anything up to 20 minutes later at a location upstream. Assuming this speed to hold true at every stage of the shockwave, the tail-end of the shockwave will take between 1–2 minutes to propagate a distance of 500m (average spacing of loop detectors on the M25).

Figure 4 shows examples of flow-breakdown on the M25. This diagram plots the one-minute average speeds of traffic against time (horizontal axis) and location (vertical axis). Traffic moves from bottom left to top right. Shockwaves can be seen in this diagram as diagonal lines (from top left to bottom right) of lighter-coloured (lower) speeds. On many days the shockwaves exhibit an apparent periodicity (in the region of 10–20 minutes between consecutive shockwaves arriving at a specific location) suggesting that some regenerative process might be at work.

The speed of propagation can also be derived as follows. If the capacity of a motorway lane is around 1800–2300 veh/hr then during flow-breakdown approximately this number of vehicles will join the back of a queue in that lane every hour. If each vehicle is assumed to occupy around 10m of road space in stop-start conditions then the queue extends on an hourly basis by 18000–23000m (or 18–23km). Hence the boundary between moving and stationary traffic moves upstream at a speed of around 18–23km/h.

When flow-breakdown does occur, the restriction of traffic flow may serve to improve conditions downstream, since the flow is now controlled. There are even cases where the temporary restriction caused by flow-breakdown allows traffic downstream to recover and start moving again! Observation of M25 traffic data as part of monitoring the Controlled Motorway suggests that when recovering from a shockwave, drivers are dispersed from one another, but eventually regroup with a headway distribution that makes them suitably primed for propagation of another shockwave.

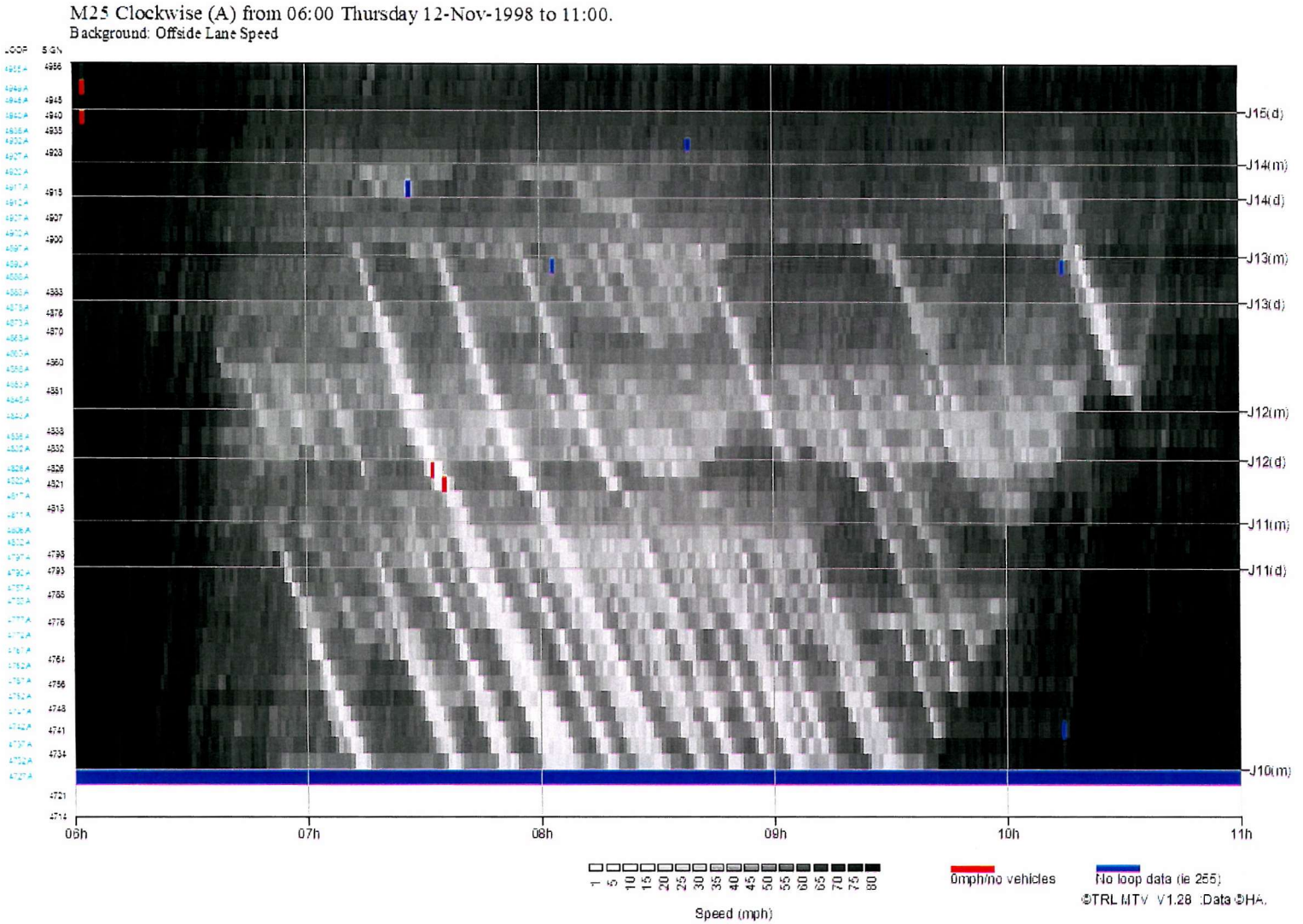


Figure 4 – An illustration of shockwaves on the M25 J10–15 Clockwise, using the Motorway Traffic Viewer

2.3 REVIEW OF LITERATURE

2.3.1 Scope of review

This review considers early theory of traffic flow, shockwaves, and some of the causes of congestion and flow instability (including some less tangible causes of flow-breakdown). Motorway junctions are a major cause of delay through congestion, and over the years a significant amount of literature has been published about this effect. A short summary of car-following models has also been included in this review because it sets the scene for shockwaves and the nature of speed-headway relationships which form an important part of the work in this thesis.

2.3.2 Early theory of traffic flow

Research work on vehicular traffic flow has followed two distinct paths, a quantitative approach using statistical methods on traffic data, and a qualitative approach using classical or intuitive mathematical models. Charlesworth (1950) described the experimental methods in use at the time, including measurement of mean speeds, flows and the differences between point measurements and journey measurements. Wardrop (1952) gave an account of the statistical theories being applied to traffic flow at that time. His focus was on the relationship between flow and density, as well as the changes in flow due to overtaking. Haight (1958) made an important contribution to the development of thinking on queue propagation.

Lighthill and Whitham (1955) took the theoretical approach in a different direction and applied the hydrodynamic modelling approach to traffic flow on crowded roads. The paper was extended to examine the movement of a high-density region along a crowded road using the 'kinematic wave model'. The emphasis was on assigning the high-density region with certain properties that would dictate its behaviour in propagation. Their proposals were surprisingly accurate, defining a shockwave as a region bounded by discontinuities in the flow rate, moving slightly slower than the mean vehicle speed and suggesting that vehicles passing through this region would have to reduce their speed suddenly on entering it. Implicit in the work was the

suggestion that flow-breakdown can arise either from a particular site acting as a bottleneck or from instability in the traffic flow itself.

Prigogine and Herman (1971) extended the ideas proposed by Lighthill and Whitham. They described concepts such as 'follow the leader' behaviour as being responsible for flow-breakdown, the dispersion and absorption of kinematic waves, and perturbation-dependent stability of flow. The theory is based on expansion of classical continuity equations and the use of statistical averages for parameters when dealing with a population of car drivers. Prigogine and Herman identified a number of shortcomings in their theory, including the fact that it did not provide any indication of the flow-density relationship within the wave. The description of the shockwave itself was also recognised as limiting because it did not allow the possibility of wave deformation over time.

One of the key assumptions implicit in the fluid-based theories is the homogeneity of the medium in which the wave is propagating. This assumption holds valid for fluids but not necessarily for traffic – where the behaviour of an individual driver is not predictable, and the distribution of vehicle types within the flow is also non-homogeneous.

May (1994) provided a review of the earliest papers on shockwave analysis. Before the 1950s only photographic techniques were being used to analyse density characteristics. May cited a study by Johnson (1928) in which aerial photography was used to observe that density along the Washington–Baltimore highway was greater at intersections and near the two cities. Another early paper by Richards (1956) began by assuming a linear speed-density relationship and then focussed on the discontinuity of density, the basic identification of a shockwave.

2.3.3 Flow-breakdown on motorways

There are two basic approaches to analysing flow-breakdown in the literature:

- traffic passing through a capacity restriction, considered as a standard queuing problem, relying on queuing theory to develop the analysis;

- traffic modelled as a fluid and wave representation, used to examine changes in traffic speeds as well as densities through the affected region.

Chin (1996) compared the two approaches. For shockwave analysis traffic was considered as a fluid, and wave representation was used to model the changes in speed and density through the 'restriction'. Chin observed that traffic wave propagation is essentially a non-linear function, but for the purposes of analysis it is often assumed to be linear. He concluded that whilst shockwave analysis provides a better understanding of the congestion situation and the physical dimensions of the moving queue, it requires assumptions about speed-flow relationships which may not be accurate.

Kulkarni, Stough, and Haynes (1996) suggested that traffic flows are self-organising. Their model implied that there is little or almost no *advance* warning of when and how congestion and flow-breakdown occur. The complexity of the problem was explained, with individual commuters leaving at different times each day, and sometimes varying their destinations. Individual commuters might also vary their driving behaviour, thus giving rise to a highly stochastic macro-level system. Speed would vary simply as a result of interaction with other vehicles; these perturbations may be enough to trigger flow-breakdown. However, the paper does not cover the possibility of specific locations combining with unstable flow as catalysts for flow-breakdown.

Kühne (1991) suggested that chaotic behaviour may be used to describe the otherwise 'irregular' pattern for shockwave occurrence. His model had two control parameters (bottleneck capacity and mean density), which determined the traffic behaviour. Regular shockwaves appear as stationary solutions of the model. At the bifurcation point the traffic will move into either supercritical or sub-critical dependence on the control parameters. These two states correspond to flow-breakdown or free-flow.

Helbing (1997) reviewed a family of models based on the continuity equation for a fluid, and suggested a new approach to account for vehicular space requirements and velocity fluctuations due to imperfect driving. In his parallel empirical study, Helbing agreed that the normal distribution was a good

approximation for the average velocity distribution. His hypothesis was that the larger the fluctuation of speed within a group of vehicles, the more likely flow-breakdown is to occur. However, in grouping the traffic data by density Helbing found that there was a "significant deviation of the empirical relations from the respective discrete normal approximations", which occurred only at traffic density of 40veh/km/lane. His explanation was that the averaging interval might have been too long due to rapid stop-and-go waves. However, having performed a linear stability analysis of his new microscopic model, he made the assertion that traffic is in fact unstable above a density of 12veh/km/lane – a value far below the 30veh/km/lane he had identified empirically as the threshold for stop-and-go waves.

Helbing also identified the tendency for driver behaviour to conform across all lanes, such that the occurrence of a shockwave on one lane is followed seconds later by similar occurrences on the other lanes.

Flow-breakdown due to roadworks is not directly related to this study, although roadworks do create their own seed-points. Hunt and Yousif (1994) identified and evaluated the major contributory causes of capacity reduction at motorway roadwork sites. Within their study period, the largest contribution to congestion came from driver behaviour at the merge area.

Edie (1963) described traffic behaviour at a tunnel. He specified a bottleneck arising from a curvature in the alignment coupled with an ascending gradient. Discontinuities in the flow profile for the tunnel occur at the base of the upgrade and again more sharply at the brow. Little research on gradient effects has been conducted, other than to identify them as potential bottlenecks. A paper similar to the work of Edie was published by Chin and May (1991). Here traffic flow is investigated in a tunnel about 200m downstream of the tunnel exit and the characteristics of the geometry are described.

Another common cause of flow-breakdown on UK motorways is the so-called 'rubbernecking' phenomenon. It should generally be considered as the effect of an incident or distraction to traffic on the opposite carriageway. Abou-Rahme *et al.* (2000) and Dixon *et al.* (2002) identified it through distinct

traffic patterns within the M25 detector data. Dudek, McCasland, and Burns (1988) described the benefits of setting up 'accident investigation sites' to deal with the aftermath of an accident out of the sight of passing motorists.

2.3.4 Effect of motorway junctions

Over the last forty years, a significant volume of research has concentrated on flow-breakdown arising in the vicinity of motorway junctions, (especially motorway merges). The relationships between slip-road flows and main carriageway flows, the capacity of the motorway near and away from junctions, the different types of merging manoeuvre, and the effects of different geometrical merge layouts have all been studied.

Chang *et al.* (1994) described an integrated real-time ramp metering model for non-recurrent congestion. Their study focussed on the real-time aspect with short execution times and the model works by comparing the traffic conditions with a linear flow-density model. The model illustrated the potential for real-time control.

Elefteriadou, Roess, and McShane (1995) looked at flow-breakdown on a merge. Their work was based on merge junctions, but moved away from identifying a maximum capacity and looked at conditions when flow-breakdown actually occurs. This is not always the result of peak volumes. The paper presented a probability function built on the occurrence and size of on-ramp vehicle 'clusters' instead of simply the volume. Elefteriadou (1997) extended the research to describe the mechanism by which flow-breakdown might occur at a merge junction based on the characteristics of the clusters identified on the ramp.

The results from a UK-wide survey containing 35 sections of three lane motorway were presented and analysed by Hounsell, Barnard, and McDonald (1994). They identified the different types of flow-breakdown, with a speed reduction followed by a flow-breakdown if the conditions were right. The condition of traffic immediately before speed breakdown – especially in the offside lane – is associated with headways of less than 1 second for speeds of around 100km/h. Mathematical approaches to identifying the location of

flow-breakdown and values for capacity were presented. The work also touched on the effects of flow composition, with the percentage of heavy goods vehicles being of particular significance. A probabilistic approach was adopted here, with a distribution for total flows in vehicles per minute. Due to the spacing between video cameras on the survey, the particular causes of flow-breakdown were not identified. However, the 'capacity funnel' and the 'merge phenomenon' were both alluded to as possible mechanisms. Their paper proposed approaches for defining both motorway capacity and merge capacity.

Hsu and Banks (1993) described a problem that is common to the M25, namely that of effects from flow-breakdown at a location downstream influencing the stability of traffic at a particular location. Most significantly, they found that the queue building back from the downstream bottleneck then affected the ability of traffic to join at the merge point. They examined the flow-occupancy relationship over a whole section of freeway and synthesised a curve looking for the effects of location on the interpretation of the data.

Banks (1991) discussed the theoretical aspect of a previous study, which examined flow processes at a freeway bottleneck. He identified that the capacity immediately downstream of the bottlenecks tended to decrease by a small amount when traffic was congested, and that the process of flow-breakdown seemed to be triggered by speed instability. The flow decrease appears to be related to the increase in vehicle passage time. Results from the more recent Controlled Motorways monitoring compare well with these observations, with shockwaves appearing in traffic data as distinct regions of relatively high occupancy.

Models have also been used to describe flow and behaviour at motorway junctions. Michaels and Fazio (1989) reported on a model of freeway merging, which was explicitly based on driver-behaviour. The model considered relative velocity and distance separation and was tested successfully on over 100 merges, providing estimates for suitable merge lane dimensions. Stanton (1992) described driver behaviour in the merge area and the subsequent flow-breakdown arising from conflict manoeuvres. Fazio

et al. (1990) developed a model of diverging on the basis of purely behavioural considerations. By examining the critical distance required to perform the diverge manoeuvre, and modelling the decision process, it was possible to assess the performance of junction design. However, the paper does not cover flow-breakdown as a result of diverge activity.

Wang *et al.* (1993) described the application of INTRAS, a microscopic simulation model which predicts both flows and lane change rates. In this model, vehicle interaction was deterministic (whereas empirical models reflect something of the stochastic nature of flow). The emphasis of the study was on determining the change in capacity brought about by merge activity. The capacity differentials highlighted in that paper can sometimes become seed-points for flow-breakdown, but an improvement in capacity does not necessarily alter the driver behaviour through the section.

Other motorway models include FREWAV (Ostrom *et al.*, 1993), WEAVSIM (Zarean and Nemeth, 1988), and SISTM (Hardman and Taylor, 1992). In each case, vehicle behaviour is governed by a car-following algorithm and a lane-changing algorithm, taking into account the individual properties of each vehicle. Each of these models has limitations (mainly in the interaction between these various algorithms) which present difficulties with the replication of shockwaves observed in traffic data.

2.3.5 Car-following models

Car-following algorithms are mathematical models describing interaction between individual vehicles in a traffic stream. The basic functional form of many microscopic models is based on a psychological argument where the behaviour of each driver can be described by the following relationship:

$$\text{Response Now} = \text{Sensitivity} \times \text{Stimulus Earlier}$$

The response can be classified as braking or acceleration, and the stimulus is the deviation from a zero relative speed. Each class of car-following model is characterised by its particular definition of sensitivity. In mathematical notation if $x_n(t)$ is the position of vehicle n at time t, then the model can be expressed as

$$\ddot{x}(t+\tau) = \alpha(t,\tau) f(\dot{x}_n(t) - \dot{x}_{n+1}(t))$$

where $\alpha(t, \tau)$ is a sensitivity function and τ is a constant time lag between successive evaluations. By using calculus, the expression can be integrated and a speed-density function derived. The stimulus $f(...)$ is usually a function of the relative speeds consecutive. The resulting car-following model would be developed with the intention of validating it using some macroscopic model, and reproducing a flow-density curve with sufficient accuracy.

Chandler *et al.* (1953) and Pipes (1967) considered sensitivity as a constant, inversely proportional to the time lag τ. This represents the simplest form of the model, and supports the macroscopic relationship between flow and density proposed by Ashton (1966) as:

$$q = q_{\max}\left(\frac{\rho_{jam}}{\rho} - 1\right)$$

Gazis, Herman and Potts (1959) proposed an inverse relationship between sensitivity and relative distance, on the basis that as vehicles moved further apart, the following vehicle became less affected by changes in the relative speeds. Gazis, Herman and Rothery (1961) went on to consider a model with a non-linear sensitivity parameter, inversely proportional to the distance headway between vehicles. Pipes demonstrated that this model corresponded to the form of macroscopic models proposed by Greenberg (1959) and Underwood (1961):

$$q = \lambda\rho \ln\left(\frac{\rho_{jam}}{\rho}\right)$$

These models do not perform well when flow is low, though it is unlikely that vehicles are following one another during such traffic conditions. Edie (1961) combined the advantages of each model by splining the Underwood model below jam density and the Greenberg model above jam density. Gazis *et al.* (1961) attempted to obtain parameters for the general form of the car-following equation. However, they could not draw any firm conclusions as the parameters were not connected to any real driver characteristics.

Pipes (1967) considered the use of perceptual factors in proposing a model of this form and Herman *et al.* (1959) proposed a model based on drivers looking two vehicles ahead. Newell (1961) considered a car-following model that described the velocity of the vehicle explicitly in terms of a velocity-headway function. The function chosen was derived from data but contained some free parameters.

Bando *et al.* (1995) suggested a similar type of model but this time described the acceleration of the vehicle explicitly in terms of a velocity-headway function,

$$\ddot{x}_n(t+T) = a\big(V(x_{n-1}(t) - x_n(t)) - v_n(t)\big)$$

where a is an acceleration constant and $V(b)$ is velocity expressed as a function of headway. The model was tested with vehicles on a circular road and converged into a series of 'dense-sparse' regions where the density was alternately high then low.

Holland (1998) provided a detailed stability analysis of models such as those described above. Stability analysis of car-following models was pursued in order to explain sudden unexpected flow-breakdown, where small disturbances to a traffic stream are amplified as the disturbance moves upstream. A stable car-following model is categorised by perturbations that are smoothed out as they propagate upstream. Holland showed that if anticipation-time is greater than reaction-time, a driver will anticipate upcoming waves and smooth them out. He also discussed how individual drivers may have a stabilising or destabilising effect on the traffic, namely that drivers stabilise flow more if their desired speed does not change significantly as headway changes. Holland concluded that flow is more stable when vehicles are spaced further apart. Mason and Woods (1997) described the stability of a platoon of vehicles and proposed a method to distinguish the impact of driver behaviour on overall stability.

Seddon (1972) and Gipps (1981) followed a slightly different route, relating reaction time to the interval between successive recalculations of acceleration, speed and displacement. Gipps in particular used parameters that corresponded to obvious driver or vehicle characteristics. The model

incorporated damping and amplification of disturbances to traffic and also allowed freedom to switch between braking and accelerating, thus replicating the *oscillatory behaviour* that characterises unstable traffic flow. Benehokal and Treiterer (1988) developed a car-following model along similar lines, incorporating traffic characteristics such as inertia and modelling the dual behaviour of traffic in congested and non-congested situations.

More recently, Addison and Low (1994) and Jarrett and Zhang (1997) used non-linear analysis for dynamical systems to identify chaotic behaviour in traffic flow using car-following models. Some replication of the oscillatory behaviour for successive vehicles was obtained through accumulation of chaotic behaviour.

To summarise for car-following models, some of the early studies involved parameters that did not correspond directly to driver or vehicle characteristics. The models of Newell and Bando are innovative but still contain free parameters. Intuitive models (such as Gipps) contain parameters more closely related to driver or vehicle characteristics. But none of these models can really replicate shockwaves without significant adjustment of the parameters. However, shockwave propagation is the best indication of the location of a seed-point and since car-following models attempt to describe this behaviour the approach is given further consideration in Chapter 4.

2.3.6 Incident detection approach

Most incident detection algorithms work by detecting the congestion caused by an incident and flagging up warnings elsewhere in the system. Gall and Hall (1989) explored the logic for distinguishing between incident congestion and recurrent congestion. Abou-Rahme *et al.* (2000) illustrated the distinction in terms of shockwave patterns within detector data.

Busch and Fellendorf (1990) empirically compared a number of well-known incident detection algorithms designed for use with traffic data. Each algorithm detects incidents by recognising unexpected changes in the traffic data, and the authors recommended using a Kalman filter along with a robust local forecast algorithm to detect significant shockwaves (see also Kalman,

1960). Nihan (1993) presented the results of a large application of this approach on the Washington State DOT ramp control system, describing the advantages of time series analysis over pattern recognition in terms of predictive capability.

Ritchie and Cheu (1993) illustrated another type of approach to the historical learning required for incident detection. An artificial neural network was trained with data obtained from a microscopic freeway simulation and then used to classify traffic at regular intervals as either incident-free or incident-related. The change in state was used to trigger an incident alert. Rose and Dia (1995) tested the approach using a similar model for detector data.

Taylor (1998) conducted a detailed review of incident-detection algorithms, with the focus being on their ability to detect incidents that lead to the build-up of a queue. The California Algorithms compare traffic conditions between adjacent detector stations. Analysis is based on variables derived from minute by minute reading for loop detector occupancy. Developments to the California Algorithm brought about a clearer distinction between incidents and shockwaves. A reproduction of California #8 applied to M25 traffic data is presented in Figure 5. The graphical display shows shockwave states in blue, whilst alarmed states are shown in red. For this particular application, the parameters required by the algorithm have been optimised to detect shockwaves.

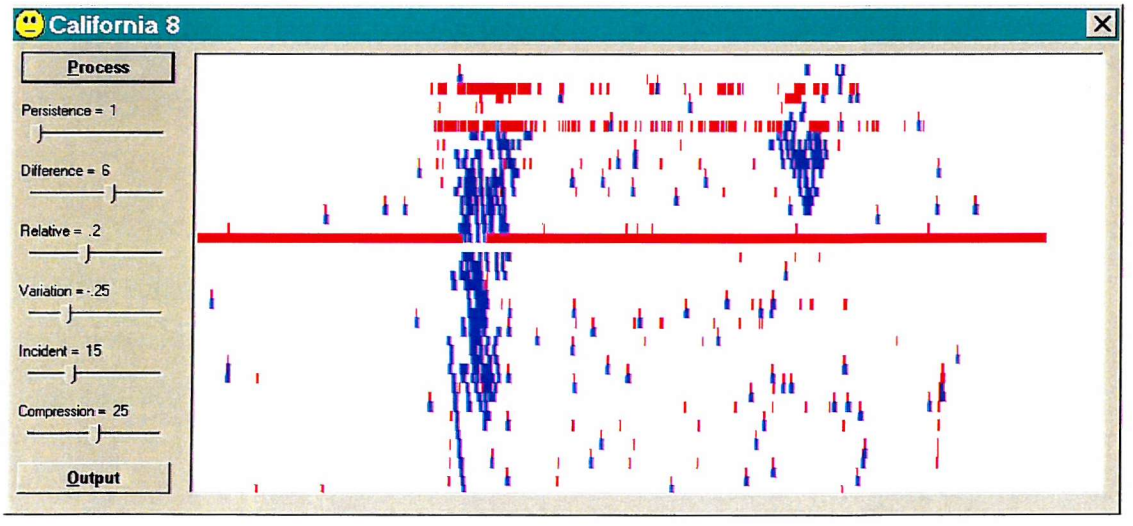


Figure 5 – A sample output from the California #8 incident detection module, showing basic shockwave patterns in blue

The McMaster Algorithm is a station-based algorithm (single point), unlike the California class that examines the differences between consecutive loops. This algorithm calibrates each location separately in order to account for the level of variation at a single site. The McMaster algorithm distinguishes three possible flow conditions based on the values of the occupancy-flow pairs, namely uncongested flows, upstream blockages and congestion. The congestion boundary can be defined using a minimum occupancy threshold. The boundary between upstream blockages and uncongested flow is defined by a quadratic function that is best calculated using uncongested data. The calibration process helps derive this by extracting occupancy-flow pairs for points that exceed a minimum speed and determining the least-squares optimised fit. The quadratic boundary is given by subtracting various multiples of the standard deviation between the curve and the selected points. This 'relative-approach' improves the transferability of the algorithm between different sensors as it eliminates the need to know the actual ranges of flow values at each individual location. Developments to the McMaster algorithm over the years (see for example Gall and Hall, 1989) have rendered its logic similar to the California #8 in that it proposes to distinguish between incident and recurrent congestion.

To summarise, incident detection can identify a shockwave once it has happened and the necessary conditions are being fulfilled, but this does not necessarily assist with locating the origin point prior to (or during) the first stage of propagation. Examination of traffic patterns to derive a predictive approach would provide more time to act in dealing with the effects of flow-breakdown, but any predictive system would need to be coupled with an effective incident detection algorithm to *validate* and evaluate the outputs.

2.4 FOCUSSING THE STUDY

In reviewing the literature for this study a number of papers were of particular interest. Hounsell, Barnard, and McDonald (1994) presented statistical analysis for a large database containing information on recurrent congestion. The role of probability in transport research work is a longstanding one; probability distributions are often assigned to a variable where discrete

decisions need to be made. However, attempting to describe the probability of flow-breakdown occurring with any accuracy, purely on the basis of evidence upstream (or prior to the event) is relatively new. Elefteriadou, Roess, and McShane (1995) described a motorway-merging model in terms of the probability of a cluster of a given size appearing on the ramp at any time.

The theoretical basis for this thesis builds upon concepts proposed by Hounsell, Barnard, and McDonald (1994) and Elefteriadou, Roess, and McShane (1995). In defining the scope of the problem there are two main points to consider, namely detecting the conditions that signify the start of a shockwave and detecting the actual propagation. Therefore, the objective of this research is to develop the probabilistic approach and apply it to traffic data in order to determine the probability of flow-breakdown occurring under specific traffic conditions. The method consists of detecting seed-points, examining traffic conditions around those seed-points, developing a probabilistic relationship for the occurrence of flow-breakdown, and using these results to make predictions.

The next chapter describes an empirical survey on the nature of seed-points. Chapter 4 presents the conceptual framework within which a probabilistic model is developed, and Chapter 5 describes the approach to seed-point detection and the composition of the database for analysis.

3 PRELIMINARY ANALYSIS OF SEED-POINTS

3.1 INTRODUCTION TO STUDY

This chapter presents the methodology and results of a preliminary study conducted by the author in order to understand the causes of congestion and determine the relative contribution of each cause to the overall congestion on the motorway.

Comprehensive monitoring of the M25 Controlled Motorway was conducted by TRL Limited for the Highways Agency, a summary is provided by Dixon, Harbord and Abou-Rahme (2002). In this context, the author carried out a more detailed investigation on congestion over a three-month period between October 1996 and January 1997. Congestion on both carriageways between J10 (A3) and J15 (M4) was monitored using detector data at 500m intervals covering both directions of the 23km section of road.

The events which lead to shockwaves occurring are also summarised below. A more detailed treatment of this study can be found in Still, Abou-Rahme and Rees (1997).

3.2 CONTEXT OF PRESENT STUDY

A simple model is presented in this section in order to provide context for the present study. Examination of sample shockwaves contained in the loop data suggested a speed drop of between 15–20km/h in the early stages of shockwave development. Once this has been flagged, it is necessary to determine whether the fluctuation in speed will work its way upstream and become a shockwave, or whether it remains a local disturbance.

If an average speed of 19km/h is applied for the shockwave (based on a simple gradient calculation using the MTV Plot), then the position of the shockwave tail can be determined over time. Assuming this speed to hold true, the shockwave will take between 1–2 minutes to propagate back around 500m to the previous detector upstream.

Based on the assumptions already discussed, the simple shockwave propagation model can be illustrated on a distance-time graph as shown in Figure 6.

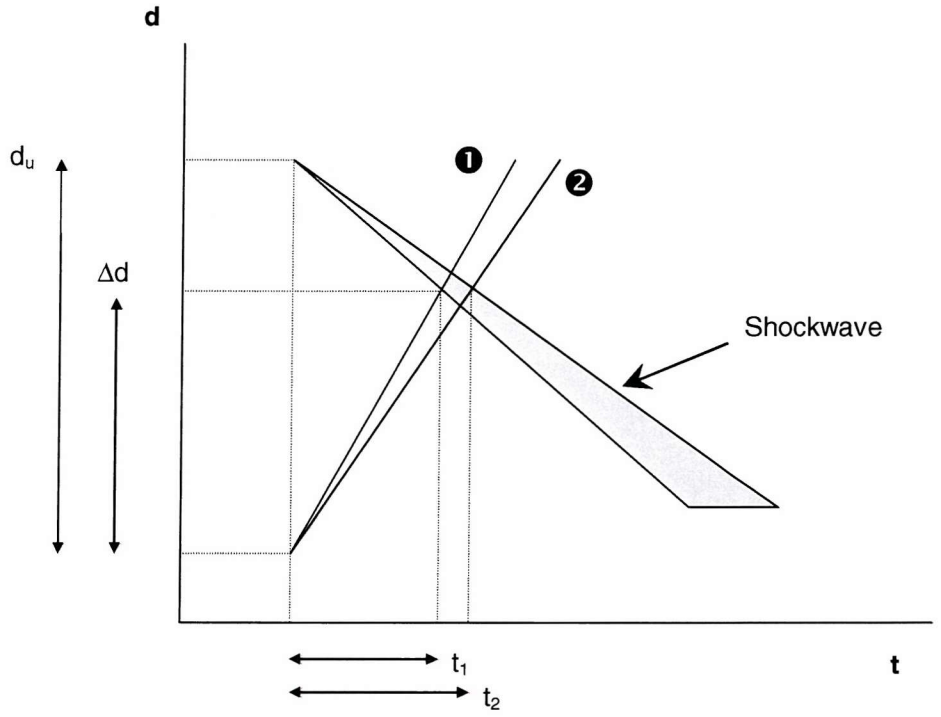


Figure 6 – Schematic representation of shockwave and vehicle trajectories

The grey triangle is a schematic representation of the shockwave propagating upstream with increasing time. The single lines show the estimated trajectory of vehicles upstream of the seed-point at the same time as the shockwave commences. When the lines cross the triangle the vehicles have encountered the shockwave, and the resultant braking will form part of its propagation.

The gradients of the trajectories are defined as follows:

❶ $v_1 = \frac{\Delta d}{t_1 - t}$ moving at current speed

❷ $v_2 = \frac{\Delta d}{t_2 - t}$ the speed needed to clear the shockwave

It can be seen that (t_2 – t_1) is the width of the shockwave. From these equations the speed-limit required to slow the upstream traffic back long enough to avoid the effect of the shockwave can be determined. The smaller

the distance between the shockwave tail and the point of effective speed control, the lower the speed-limit required to achieve the outcome.

The relationship between Δd and the initial distance between the seed-point and the vehicles upstream is linear. However, the relationship between the required speed-limit and the duration of the shockwave effect is inversely proportional. Figure 7 shows three sample profiles for a platoon of vehicles detected 1.5km, 3km, and 6km upstream of the seed-point and travelling at 50mph.

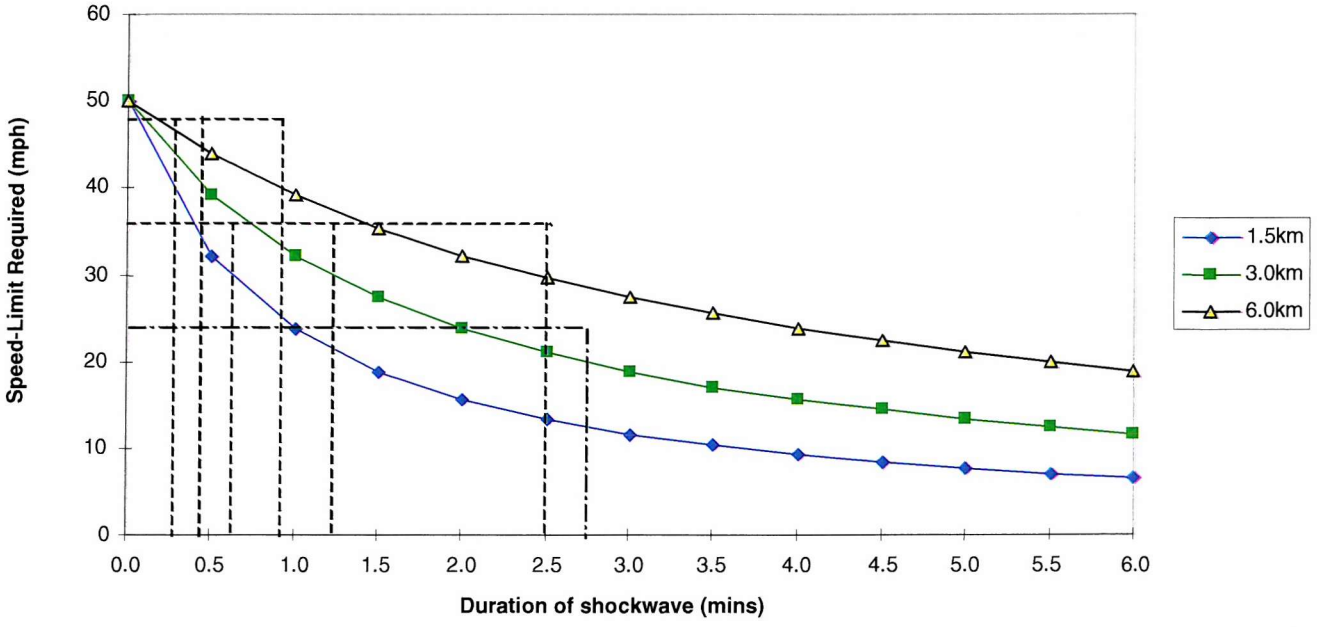


Figure 7 – Plot of required speed-limit against duration of passing shockwave

For the distances involved along the Controlled Motorway section of the M25 it can be seen that speed-limit settings of 50mph and 40mph will provide little extra time for vehicles 3km upstream to accommodate the passing of a shockwave. According to this simple model, a 30mph speed-limit will provide a space of 2.5 minutes when displayed to vehicles 6km upstream and 1.5 minutes when displayed to vehicles 3km upstream of the site where the shockwave is first detected.

However, lower speed-limits do have an important effect in this model. The application of a 20mph speed-limit 6km upstream of a seed location will

provide a space of nearly 6 minutes, long enough for downstream traffic to recover from a severe shockwave. Even at 3km upstream the 20mph setting provides a space of nearly 3 minutes. These speed-limits might be displayed to drivers as part of a sequence, and during periods of congestion when flow-breakdown has already taken place, the overall effect on the motorway could be beneficial.

It is here that the present study finds application. The ability to predict the occurrence of shockwaves (or at least describe the probability of occurrence with some degree of confidence) would significantly assist in the implementation of control systems. In terms of the simplistic model described above it is clear that a time contingency would allow control using acceptable speed-limits (such as 40mph) making shockwave dissipation a more feasible function of the control environment.

3.3 THE ORIGIN OF SHOCKWAVES

3.3.1 Defining a seed-point

A seed-point is the location at which a shockwave begins, and from which the propagation of a shockwave can be traced. This origin point can be considered as a combination of volatile traffic conditions, the effect of local geometry, and a specific disturbance. A sudden change of lane in busy traffic, or a surge of demand through a curved section of motorway would both be candidate events for the commencement of a shockwave.

Whilst a significant amount of information is available regarding shockwaves and their propagation, relatively little is known about the characteristics of seed-points. We can consider a seed-point as an unknown process that receives inputs (flows coming from upstream), and generates outputs (either stable flows or shockwaves). A shockwave may begin at a specific location but its effect does not remain in the same location. Occasionally the origin point itself can be moving, as in the case of an escorted heavy load.

Preliminary examination of video footage and corresponding traffic data from loop detectors shows that causes of congestion can be classified under one of

three generic headings ('unpredictable', 'predictable' and 'recurrent' congestion).

3.3.2 Unpredictable congestion

Unpredictable congestion covers those situations where it is not possible to predict the time or location of an event that could potentially cause congestion. Figure 8 shows some examples of these unpredictable causes, including accidents, animals on the carriageway, vehicle breakdown or combustion, distractions such as low altitude aeroplanes, security alerts, shed loads and extreme weather conditions.

3.3.3 Predictable congestion

Predictable causes of congestion include those situations where it may be possible to predict the time and location of an event that has the potential to cause the congestion. Figure 9 shows four examples of such causes namely low sunlight, fixed lane closures, mobile lane closures, and the passage of an escorted slow-moving vehicle. A certain amount of preparation might be possible here including traffic alerts and police presence, but in general the ensuing congestion is simply accepted.

3.3.4 Recurrent congestion

When traffic flows are lower than the nominal capacity of a carriageway or junction, the traffic will flow without any significant reduction in speed. As traffic flow approaches the capacity of the carriageway or junction so the probability of congestion occurring rises. Congestion here is defined as slow-moving high-density traffic often characterised by speeds varying from 0–30mph. Because this problem mainly occurs at high traffic flows the resulting congestion can frequently be extensive. Figure 10 illustrates the three main locations along a motorway where congestion caused by weight of traffic can occur, namely at the merge, diverge, and between junctions.

Problems at merge junctions can arise from direct conflict at the merge point between those vehicles attempting to join the main carriageway and those vehicles already travelling on it.

Figure 8 – Unpredictable causes of congestion

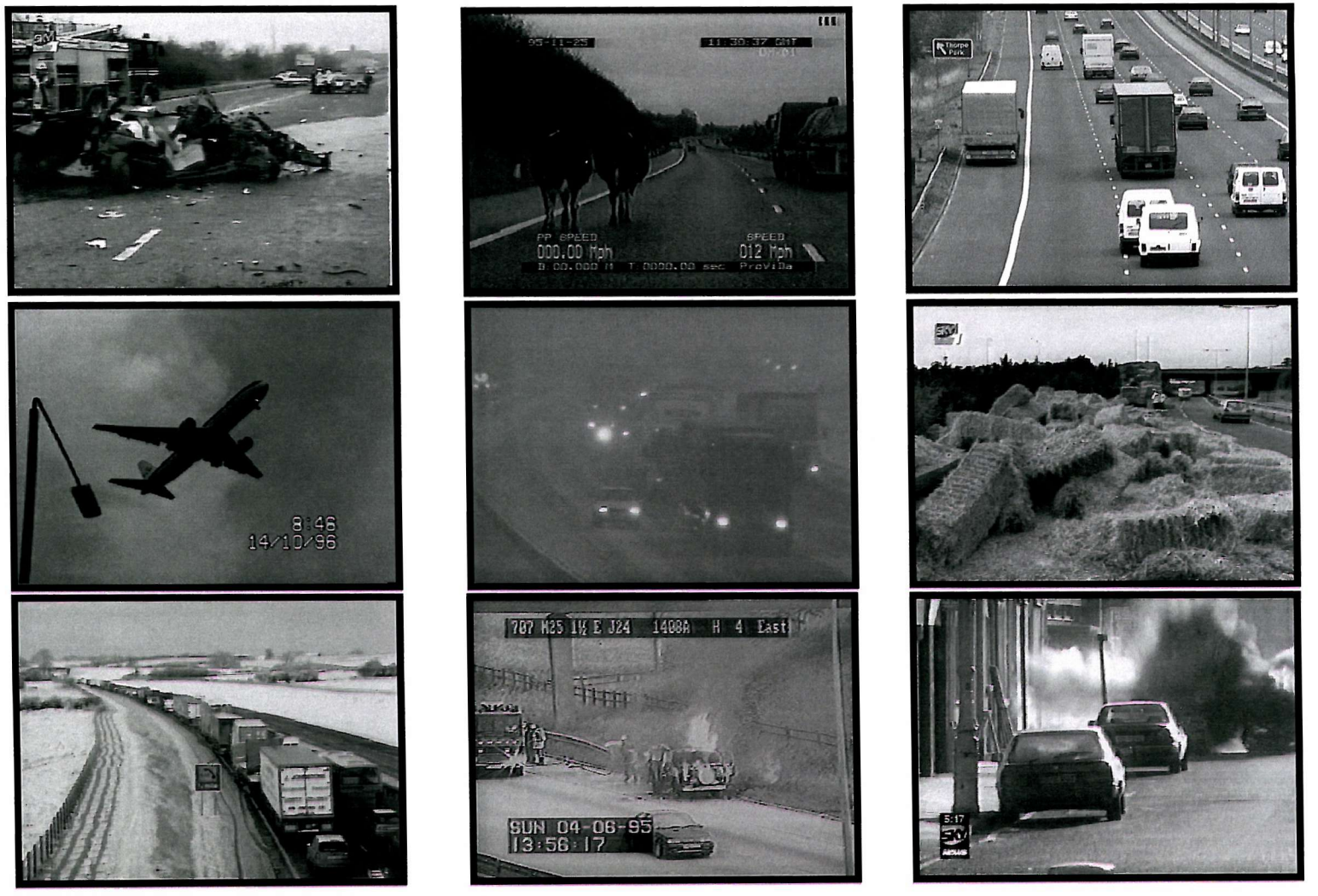


Figure 9 – Predictable causes of congestion

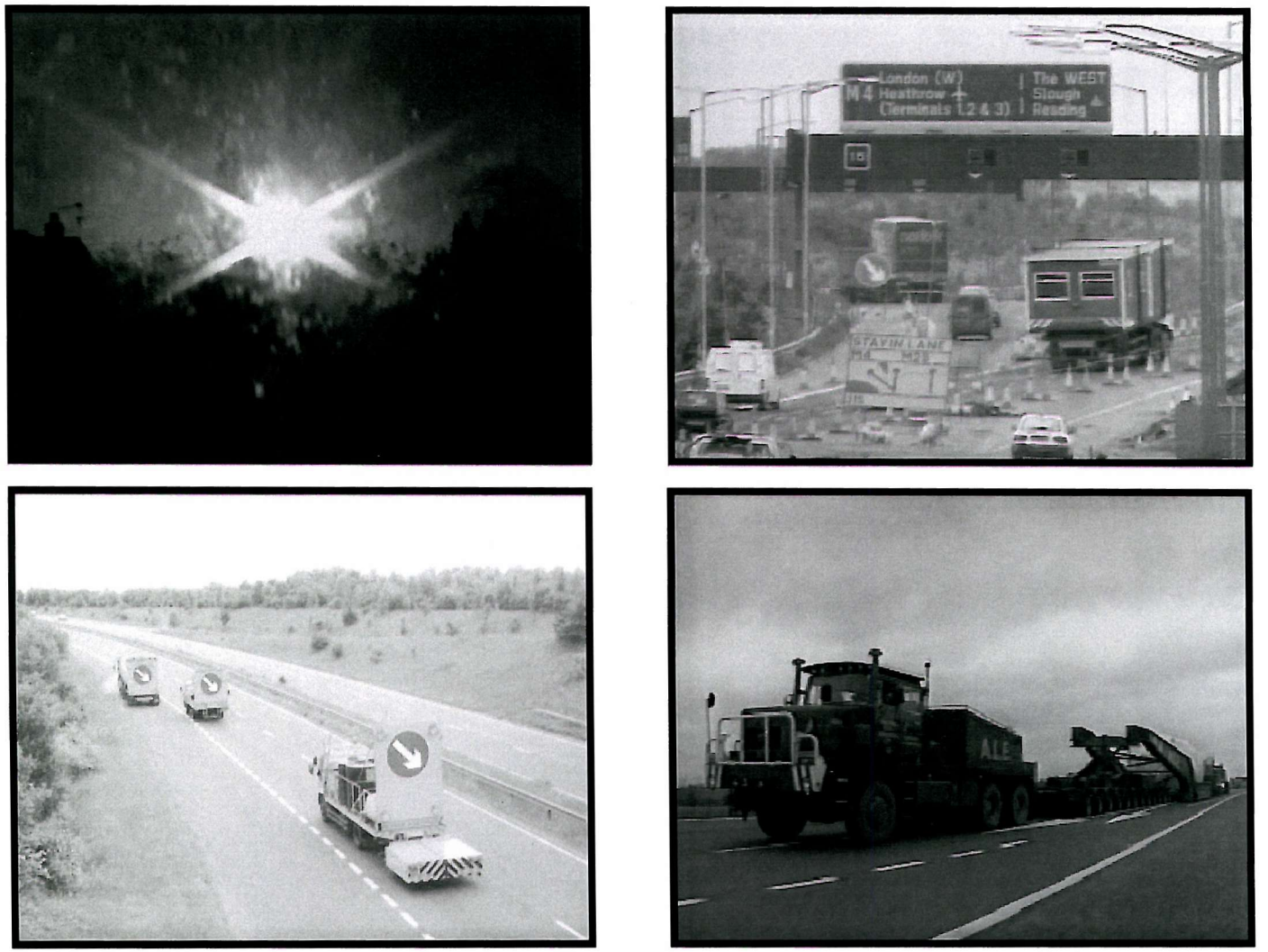


Figure 10 – Recurrent congestion

These conflicts often come about when the main flow is dense enough to reduce the number of gaps perceived as acceptable for merging drivers. In other cases, drivers on the main carriageway migrate towards the offside in order to accommodate heavy merge-arm flows. There is also a type of flow-breakdown initiated by an unsustainable increase in capacity due to temporary acceptance of a smaller headway. This capacity bottleneck tends to occur between 800–1000m downstream of the merge point.

At the diverge junction itself, a common cause for disturbance to traffic is the activity of drivers making a late lane change (the so-called 'swooping' manoeuvre). Other manoeuvres involve slowing in order to make a normal lane change, hesitation that leads to lane-straddling, and drivers coming back out of the diverge lane in order to overtake a vehicle before making their exit. It is also common for flow-breakdown to occur simply because of blocking back from the roundabout at the end of the slip-road.

Away from the junctions the causes for flow-breakdown are harder to identify, nevertheless they are there. Detailed observations of traffic behaviour indicate that the phenomenon includes the most subtle of perturbations to unstable flows, critical headway being encroached, excessive lane changing, influence of geometry and other local effects.

3.4 DATA ANALYSIS FOR M25

3.4.1 Data preparation

The loop detectors provided good quality information at one-minute intervals on average speeds and flows. This information was used to identify periods of congestion and to quantify the extent of the congestion. Data was processed from large binary files in order to extract the speed and flow information at each loop position on both carriageways. These files were then loaded onto a workstation running dedicated software and graphical packages. The Highways Agency 'Roadworks Bulletin' and the MIDAS Sign Settings database were examined and a log was kept of the times and locations of any roadworks on the M25 J10–J15.

3.4.2 Analysis using MTV software

For each day and for each carriageway direction, an overview plot of the traffic speeds on the M25 between Junctions 10 and 15 was produced using MTV. Figure 11 shows a typical pattern for the clockwise carriageway covering all day 30 October 1996 (a sample day from that study). A grey scale is used on the plots to represent the average speed. The lighter the colour the lower the average speed – with black representing 80mph and white representing 0mph. If the 0mph reading is associated with zero flow, then the white shading is replaced with red.

The early-morning period (from midnight to around 05:00) was ignored because of the large number of red areas (implying that very few vehicles are on the motorway at this time). As a tool, this overview plot was used to identify the times and location of areas where traffic speed falls significantly below the speed-limit. For example, Figure 11 would suggest that the period from around 07:00 to 11:00 exhibits a level of congestion that needs investigation. A similar plot was produced for the corresponding period on the anticlockwise carriageway.

After the initial assessment of time and location for slow-moving traffic, plots containing more detail for the area of interest were produced, again using MTV. Figure 12 shows the detailed plot for the morning peak period between 06:00 and 11:00.

With this level of detail it was possible to break down the times and locations of slow-moving traffic into a number of separate distinct zones and to allocate a unique identification number to each zone (described further in the next section). The characteristic patterns of slow moving traffic within each zone normally provide a clear indication of whether the congestion is caused by weight of traffic (i.e. recurrent congestion) or by an incident.

In Figure 12 the slow-moving traffic between 07:00 and 08:30 is caused by flow-breakdown due to weight of traffic between Junctions 13 and 14, while the slow-moving traffic from 08:30 to 11:00 shows all the characteristics of an incident.

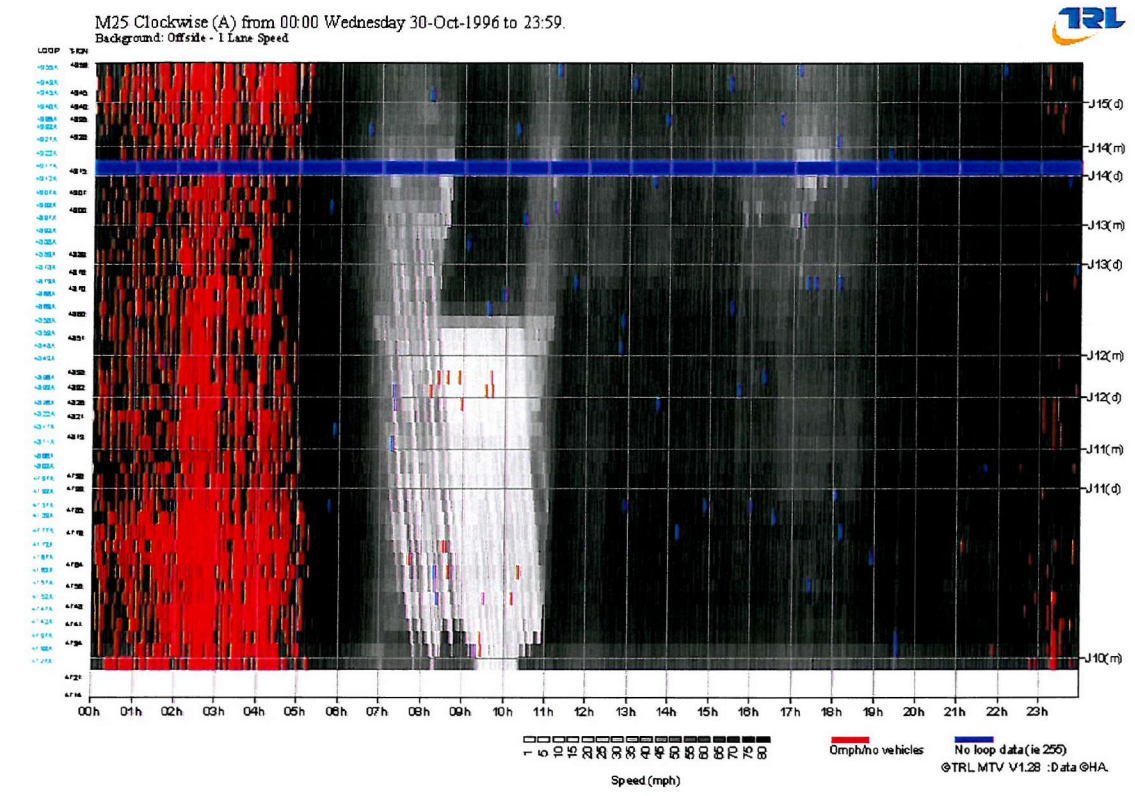


Figure 11 – MTV Speed Plot showing congestion pattern for the clockwise carriageway (30 October 1996)

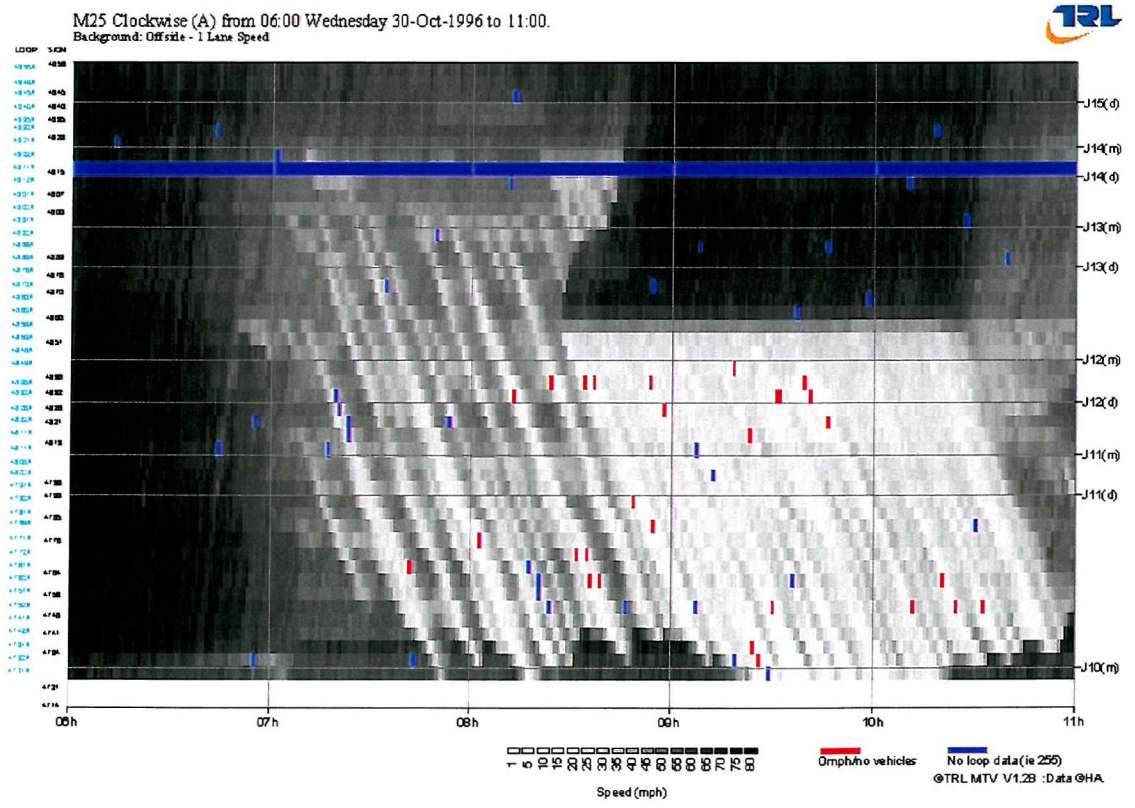


Figure 12 – MTV Speed Plot showing more detail for 06:00 to 11:00 (30 October 1996)

The traffic speeds are consistently low as the vehicles are queuing behind the incident (as opposed to the stop-start behaviour during normal congestion). Downstream the speeds are consistently high as traffic leaves the scene. In cases where the distinction is not as clear as that shown in Figure 12 the police logs help determine whether or not it is an incident.

3.4.3 Evaluation of congestion

Due to the widespread availability of speed and flow information from the MIDAS loop detectors, the 'Vehicle Hour Delay' (VHD) indicator was chosen to measure the extent of congestion in each of the zones identified on the plots. VHD is derived from a simple manipulation of the relationship between speed, time, and distance. For a defined length D with N drivers passing through in one minute the VHD for that minute is

$$VHD = \left(\frac{1}{V_{actual}} - \frac{1}{V_{desired}} \right) \times D \times N$$

It indicates the total delay (in hours), incurred by all drivers in the defined zone through being unable to maintain a predetermined desirable speed (for example 80km/h). During periods of signal control on the M25, this value represents the maximum speed for free flowing traffic.

The area of congestion is approximated by a trapezium, the four points of which are uniquely identified by coordinates of the form (time, location). Since speeds above the predetermined desired value are ignored by the VHD calculation it was not necessary to make the initial approximation too accurate! Figure 13 provides an example of this process, with two areas of congestion marked out by trapeziums for illustration purposes. The coordinates were listed in a clockwise manner beginning at the top left corner (the origin point).

The four coordinates were then processed using another purpose written tool developed by the author within MTV, designed to calculate the VHD figure for vehicles within the zone. Figure 14 shows a screenshot with sample input to obtain the total VHD for the day.

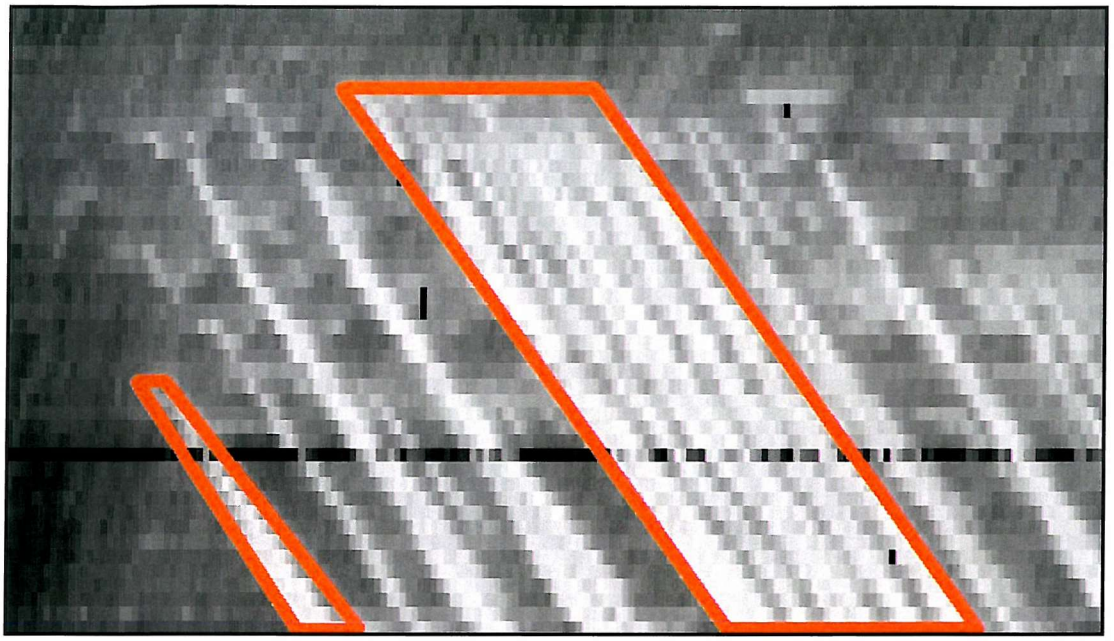


Figure 13 – Illustration of the trapezium approximation to congestion arising from two identifiable origin points

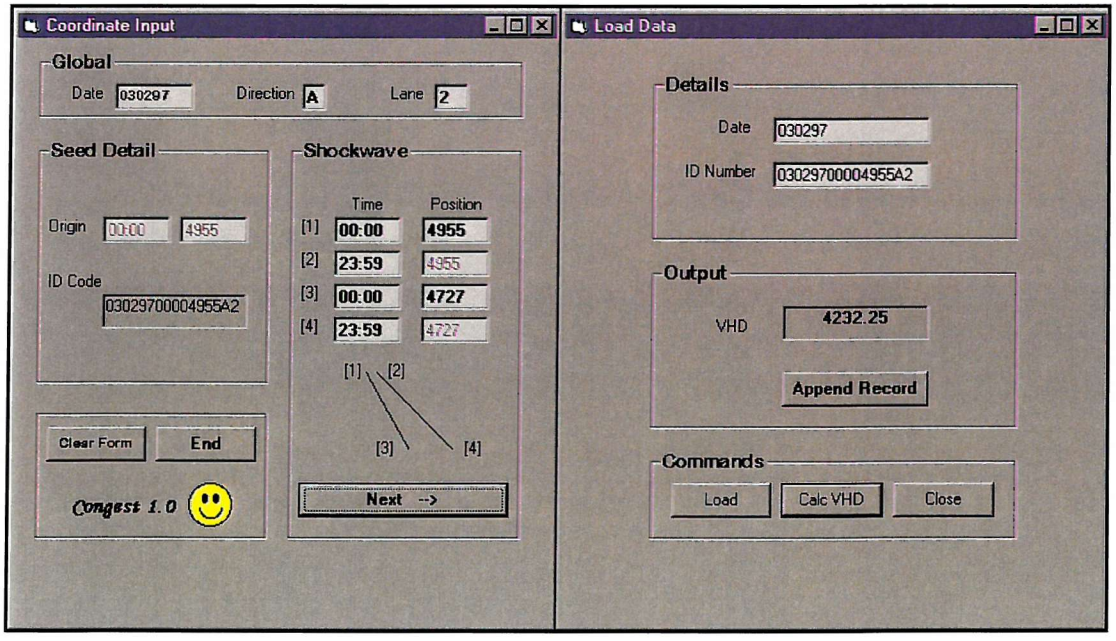


Figure 14 – Screen shot from VHD Calculation Software

The result for each zone was appended to a data file for the particular day on which that congestion had occurred. All the files were then imported directly to the final database in order to minimise the risk of data input error. Each record was assigned a unique ID Code, the coordinates of the zone and the value for VHD obtained through calculation. Implicitly contained within the ID Code for each region of congestion were the date, time and location of origin in terms of carriageway and loop site. As an overall check the VHD value was calculated for the whole day and compared with the sum of the VHD values obtained for each specified zone.

On average, the sum of individual zones formed around 90% of the total VHD for the clockwise carriageway and between 80–85% of the total VHD for the anti-clockwise carriageway.

3.4.4 Video survey

A gantry-mounted configuration was chosen for a video survey during late 1996, consisting of two cameras facing upstream and two cameras facing downstream. Cameras were discretely hidden on the gantry so as not to distract drivers. Filming downstream captures brake light sequences which can be informative with regard to shockwave triggers. Filming upstream gives a clearer view for lane changing and allows some estimate of headway to be made in certain situations. Suitable filming periods were identified from the traffic data, and video clocks were synchronised so that a comparison between the video recordings and the traffic data would be possible.

Three particular sites on the M25 were chosen for viewing on the basis that they provided the best examples of recurrent congestion caused by merge, diverge, and mid-section effects.

- M25 J13–J14 Clockwise, at seed-point between the junctions.
- M25 J11 Clockwise, downstream of the merge point.
- M25 J13 Clockwise, at the J13 diverge point.

3.5 VIDEO SURVEY RESULTS

3.5.1 Mid-section flow-breakdown (M25 J13–J14)

Flow-breakdown occurring away from the proximity of junctions can arise from a number of contributing factors. From the video survey, one significant factor appeared to be high density of traffic combined with the reaction by drivers to brake lights ahead. On this section of motorway, changes to both gradient and horizontal alignment also appear to adversely affect the behaviour of traffic, leading to flow-breakdown in conditions of high volume and high occupancy. Lane-changing manoeuvres were easily accommodated with no evidence of drivers braking as a result.

Test drives were made by the author along the clockwise section between J13 and J14 to observe behaviour and local conditions. After the merge at J13 the motorway follows a downhill gradient past marker 4896A. The uphill climb then begins at marker 4897A (100m further along in the direction of traffic flow) and some acceleration was necessary to maintain speed. However, drivers also had to anticipate how the vehicle in front would respond to the change in gradient, and so headway became shorter or longer depending on the accuracy of this judgement. Headway distribution within a cluster of proximate vehicles would have been significantly altered as drivers adjusted their speeds relative to one another. This condition continued past the loop detectors at 4900A and 4902A. Evidence for this behaviour was found in the occurrence of much frequent braking (drivers touching the brakes) with no apparent change in speed.

Between 4904A and 4906A the gradient levelled off, still uphill but less steep, and by 4907A the gradient was not noticeable. Here a second readjustment took place as some drivers found themselves going too fast. Thus there were two noticeable incidents of braking along the section observed, each of which, in conditions of sufficiently high occupancy, could become a trigger for flow-breakdown. The significant change in traffic conditions between loop 4902A and loop 4907A is therefore most likely brought about by the change in traffic state due to the gradient.

It has been observed that the capacity of a motorway merge is often found to be greater than the capacity of the main carriageway, (see Hounsell *et al.*, 1992). Because drivers accept and maintain a shorter headway in order to accommodate merging manoeuvres during peak hours, a region of high-occupancy flow is established. Evidence suggests that as this traffic moves downstream, away from the merge, maintaining a short headway is no longer necessary. This increase in headway creates a capacity bottleneck and flow-breakdown soon follows. Flow-breakdown can be attributed to this cause whenever the seed-point is located closer to J13 (upstream of 4900A).

Perturbations in the traffic flow also occur on the second curve of the s-bend between J13 and J14. The high occupancy region of traffic does not always degenerate into a capacity bottleneck, as described above. Due to the relatively short distance between junctions along that section of motorway, drivers do not always 'relax' their headway in the same way as if they were on other motorways. A distribution of short headways in the traffic flow is maintained until vehicles encounter the second bend (between markers 4906 and 4907). From the video footage it can be seen that headway distribution is then affected by an *involuntary* relaxation of headway due to drivers negotiating the bend in their own particular manner.

3.5.2 Merge-related flow-breakdown (M25 J11)

Video evidence showed that during the morning peak hour, the origin of flow-breakdown appeared to alternate between the 'capacity bottleneck' and the first significant bend. Occasionally, the flow-breakdown was not caused by local effects, but was the result of a shockwave which had originated much further downstream. It was interesting that a conflict at the merge caused perturbations to the flow but did not bring traffic to a standstill, whereas flow-breakdown arising from the 'merge capacity bottleneck' caused standstill in the nearside lanes. The effect of a shockwave coming from further downstream was distinguishable by flow-breakdown occurring in the outside lanes first, (rather than the nearside).

Design of the merge junction plays a critical part in its performance. Filming at this location identified an unnecessary conflict brought about by the

two-lane slip road merging with the three-lane carriageway to become a four-lane carriageway. Those vehicles in the right-hand merge lane were able to migrate onto the main carriageway if they so wished. A good proportion remained in lane and these vehicles were forced to merge with vehicles from the left merge lane as that particular lane came to an end. This is important in terms of what the drivers are expecting at a merge. Those vehicles on the main carriageway would anticipate vehicles merging from the left. Those on the merge arm would anticipate having to merge with vehicles from their right, although at the point where the left merge lane comes to an end, drivers were being forced to carry out a *second merging manoeuvre*, this time with vehicles from the left. Occasionally in the video, a driver was unable to merge and ended up on the hard shoulder, eventually becoming desperate and forcing an entry.

3.5.3 Diverge related flow-breakdown (M25 J13)

The videos revealed that flow on this particular section of motorway is characterised by lane changing. Aggressive or absent-minded drivers can perform sudden lane changes across more than one lane (known as 'swooping') causing other drivers to brake suddenly. As before, the effects of these perturbations can be manifested as shockwaves.

There were a few examples of drivers finding themselves in the slip road and attempting to rejoin the main carriageway after the diverge lane markings have changed. This manoeuvre was not always expected by drivers on the main carriageway and caused some disturbance.

Other possible causes of flow-breakdown include local traffic problems causing tailbacks on the slip road and the effects on the main arm flow from 'queue jumpers' who wait till the last possible instant before leaving the motorway. Such vehicles could be seen braking whilst still on the main carriageway, the driver's behaviour being influenced by what was happening on the slip road as opposed to the vehicles immediately ahead. Wedlock, Peirce and Wall (2001) studied various schemes aimed at combating the swooping behaviour in diverge areas, and demonstrated the benefits of ghost-island layouts.

3.6 RESULTS FROM THE TRAFFIC DATA

3.6.1 Presentation of results

After filtering the data, the completed database consisted of 880 acceptable instances of congestion, the origin and magnitude of which could then be determined. This equated to 66 days over the three-month period, including weekends. A complete incident log was also collected, with details of accidents deemed to be responsible for causing delay on the M25 between J10 and J15 in both carriageway directions.

The vehicle hour delay was only calculated within the Controlled Motorway section. This approach may have underestimated effects of particular causes outside this section of motorway by anything up to 15% on days without major incident. However, this was not immediately relevant for analysis being conducted within a defined area. Some rationalisation of category headings has been applied for the smallest values, for example 'tyre debris' has been amalgamated with the category 'debris' because their contribution to the overall delay index is so small.

If an incident appears to have occurred outside the study section, such that congestion arising from that incident has propagated back but the cause is not identifiable, the cause of this visible congestion is classified as 'not known'. Where a definite cause could be identified, the VHD applicable to J10–J15 was included.

Delay from accidents contributed around 19% of the overall delay for the three-month period under study. Table 1 shows the relative contribution from the 79 incidents of varying severity. The most severe accident of the period occurred on 21 October 1996, contributing 6525 out of 7271 VHD on the clockwise carriageway, and 4165 out of 4547 VHD on the anticlockwise carriageway – a total of 10690 VHD for the one accident! A heavy goods vehicle crashed through the central reservation from B to A carriageway, obstructing both carriageways for around five hours. Removing this single accident from the calculation would change the proportion of delays attributable to accidents from 19% to 15% of the new total.

Number of Accidents	VHD	% of Total VHD for Accidents
1	10,690	24
5	17,440	40
10	25,088	57
38	39,616	90
79	44,120	100

Table 1 – Contribution of individual accidents to total 'accident-related VHD' between October 1996 and January 1997 on M25 J10–J15

Figure 15 shows the split between recurrent and incident-based congestion for all identifiable VHD. It can be seen that the recurrent congestion contributes almost three times as much as the incident-based congestion on this section of the M25. In numerical terms, this amounted to some 170,000 VHD for recurrent and some 49,000 VHD for incident-based congestion over the whole monitoring period. Figure 16 shows the relative VHD on each carriageway.

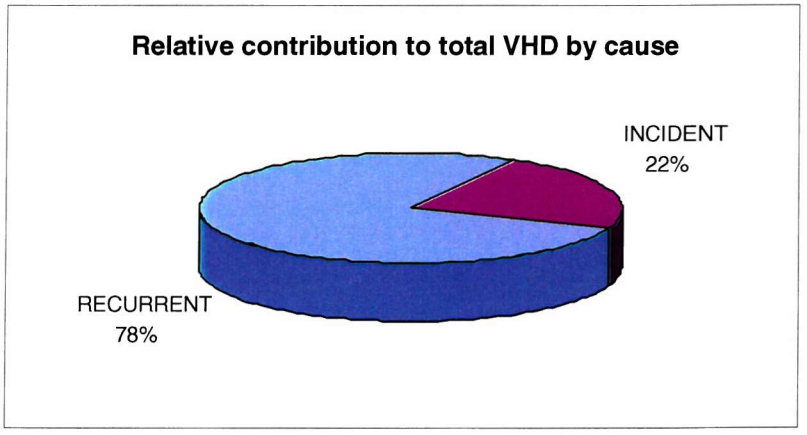


Figure 15 – Distribution of VHD between recurrent and incident-based congestion for M25 J10–J15 (October 1996 to January 1997)

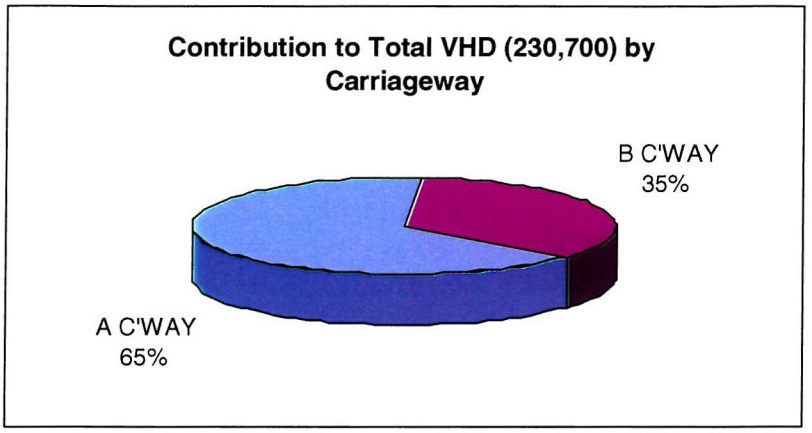


Figure 16 – Distribution of VHD between carriageways for M25 J10–J15 (October 1996 to January 1997)

Figure 17 shows the contribution of each category to the total VHD calculated on the motorway, for the 66 suitable days. The category named 'other' is broken down into further subdivisions, as detailed in the magnification of the 2%.

Both diagrams illustrate the variety of causes of congestion, as well as the relatively small contribution made by these causes compared to the effects of an accident during peak hour, or the effects of recurrent congestion. Only one definite case of 'rubbernecking' (vehicles slowing to observe an incident on the opposite carriageway) was identified during the monitoring period, although other instances have been observed (see Dixon, Harbord, and Abou-Rahme, 2002)

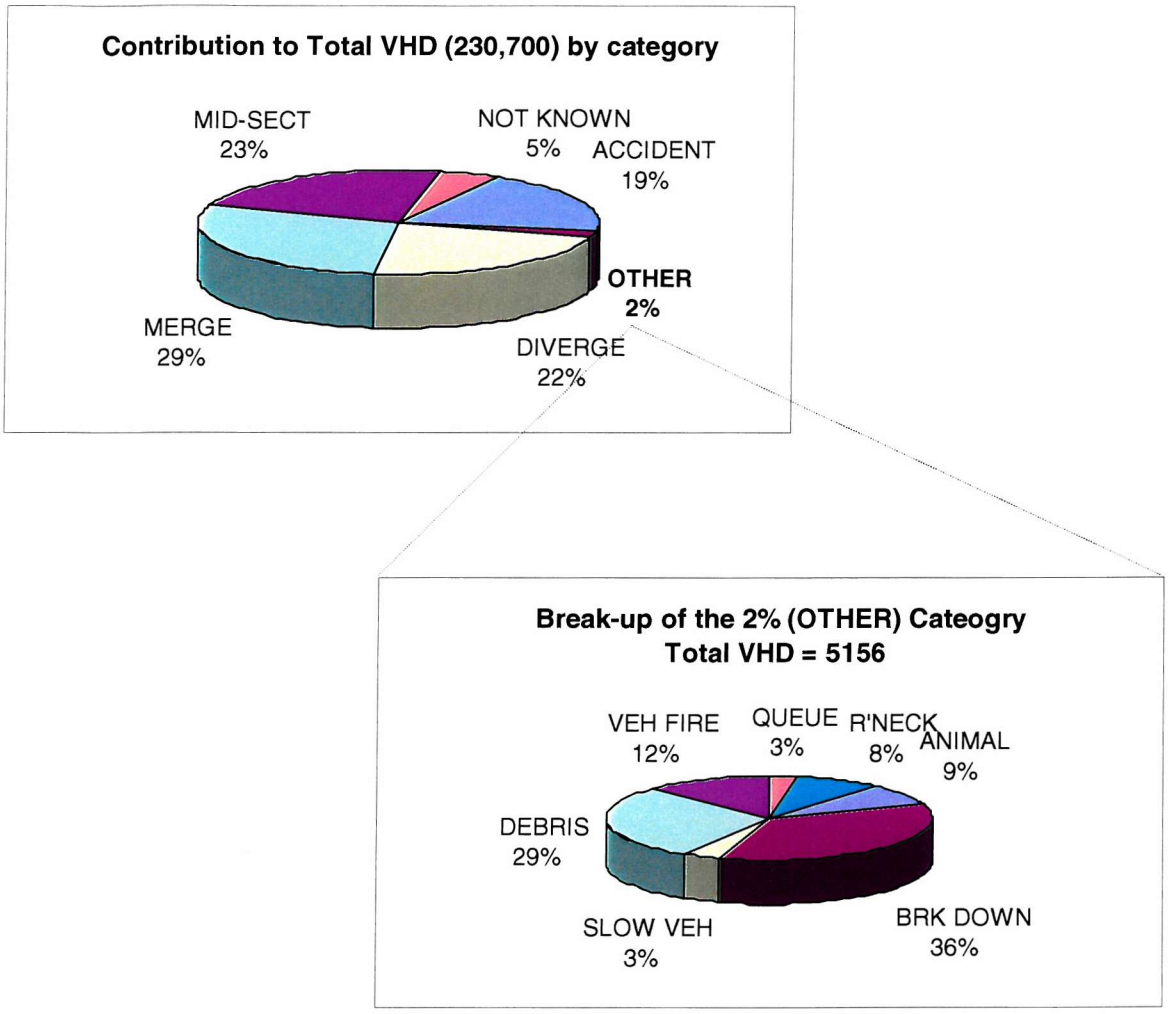


Figure 17 – Contribution to total VHD by category for M25 J10–J15, both directions between October 1996 and January 1997

3.6.2 Other causes of congestion

Whilst no security alerts were recorded on the M25 during the period of monitoring for this project, a significant security alert did occur on both the M1 and the M6, between 3–4 April 1997. The alert led to the M1 being closed for 9 hours between J17 and J19, and the M6 being closed 30 hours in the same region. The total closure brought about considerable delays and demonstrated the relative seriousness of this cause of congestion. Given that public safety cannot be compromised, the motorway must remain closed until the area has been thoroughly searched.

There were a number of instances of shed loads on the M25, but all occurred outside of peak periods and therefore did not generate any noticeable

congestion. On other sections of the M25, spillage has caused significant delays. An example from the data can be found on 06 March 1997 where a diesel spill had J2 of the M25 closed on the clockwise carriageway during the afternoon rush hour. Within the results for this study, one instance was recorded where a driver collided with the back of a queue.

The M25 runs north-south along the particular section being monitored for this project, and no significant effects of sunlight were noted. Anecdotal evidence suggests that low sunlight is a significant problem on the M40.

3.7 CONCLUSION FROM PRELIMINARY STUDY

A detailed study of congestion caused by weight of traffic was carried out on the M25 for data collected between October 1996 and January 1997. Video surveys carried out between late October and early November provided focus on both driver behaviour and possible mechanisms leading to flow-breakdown. Three generic types of flow-breakdown were investigated and conclusions drawn. In addition to recurrent congestion, a log of all congestion caused by incidents was maintained, with the causes being determined from police records and other sources of information.

Flow-breakdown in the mid-section between junctions is seemingly unrelated to merge or diverge-related problems. This study has suggested that flow-breakdown is caused by a gradient phenomenon whereby the interaction of vehicles with each other is affected by changes in *vertical* alignment of the road, leading to braking and increased instability. Changes in the horizontal alignment also appear to provide perturbations to unstable flow. Therefore, flow-breakdown in the mid-section is the result of a combination of two factors:

- the existing characteristics of the flow through that section (in terms of stability, headway distribution and quantity);
- the physical features of the section requiring any action on the part of drivers to negotiate (for example, an uphill gradient).

In the vicinity of a merge junction, flow-breakdown often occurs because of the difference in the capacity of the merge and the capacity of the downstream section. Other causes of flow-breakdown include direct conflict at the merge point itself, and a migration of drivers from the nearside lanes to the offside lanes of the main carriageway in anticipation of the merge. On the M25, the most common cause of merge-related flow-breakdown (observed in this study) is the downstream capacity problem.

From the video evidence, flow-breakdown at the Junction 13 diverge was more directly related to lane changing, especially with drivers 'swooping' across more than one lane at a time and drivers joining the diverge lane late in order to avoid queuing. Occasionally problems would arise when a vehicle in the diverge lane attempted to rejoin the main carriageway. Each of these activities causes braking on the part of other drivers, which in unstable flow conditions will lead to flow-breakdown. However, traffic blocking back on the slip-road from the roundabout at J14 made a much larger contribution to the total VHD in this study.

A number of different types of accident were responsible for flow-breakdown throughout this period, and between them contributed to 22% of the total VHD.

4 MODELLING SEED-POINTS

4.1 INTRODUCTION TO MODELLING SEEDS

Having gained a better understanding of what characterises the origin of a shockwave on the motorway, the research then focussed on how to model these seed-points.

In raw traffic data the seed-point is abstract, there is evidence that it has occurred but classifying it and distinguishing it from other traffic patterns is not straightforward. In the preceding chapters a survey of real seed-points on the M25 pointed to a number of possible causes, ranging from gradients through to behaviour at junctions. The *combination* of geometric properties and traffic characteristics was presented as the essential ingredient for shockwaves to occur.

This chapter explores two theoretical frameworks from which a seed-point can be understood. Each framework is based upon observations and research conducted by the author during his employment with TRL Limited. At the time of writing there remains much scope for development in each area.

The first of these frameworks, the '**Mechanistic Approach**', seeks to identify the origin of a shockwave through implementation of mathematical models describing car-following behaviour. The question we ask is whether there is enough scope within the car-following *mechanism* to generate shockwaves? If not then are there other means by which we might consistently induce them?

The second of these frameworks, the '**Probabilistic Approach**', is the framework within which the remainder of the current thesis has been developed. Using inferential statistics, the objective of this method was to derive a posterior distribution for the *likelihood* of shockwave occurrence, based on real traffic data.

Each of these areas is now examined in more detail.

4.2 THE MECHANISTIC APPROACH

4.2.1 A particular example of car-following

The Mechanistic Approach has to do with the generation of shockwaves through the mechanism of the car-following algorithm.

Section 2.3.5 gave an overview of different schools of thought for modelling car-following behaviour. The background to the work presented in this section comes from an investigation by Abou-Rahme and White (1999) in reinterpreting the non-conventional car-following models proposed by Gipps (1981) to explore its ability to model the propagation of shockwaves. Aspects of that work are reported here both to highlight potential within the mechanistic approach and to provide an alternative approach to the probabilistic model.

As mentioned in the literature review, a microscopic model called SISTM has been developed at TRL Limited for the Highways Agency. SISTM describes the behaviour of vehicles on a multilane carriageway, and has been calibrated over many years using traffic data collected from a number of motorways in the UK. An initial investigation by the author into the ability of SISTM to replicate speed-flow behaviour on the M25 found that whilst congestion patterns for stationary traffic were reproduced well, there was no scope for generating and reproducing a shockwave through the normal modelling process, (Abou-Rahme *et al.*, 1996).

The investigation then focussed on a particular implementation of the Gipps Algorithm within SISTM, to understand how shockwaves were being treated. The stability and sensitivity of the car-following model to its constituent parameters were examined by developing vehicle classes whose behaviour was regulated by various aspects of the car-following theory. The parameters in the model were investigated in order to permit the propagation of shockwaves following a given disturbance. The investigation of seed-points was based on the possibility that minor disturbances within unstable flow regimes may occur through ordinary vehicle interactions. In fact this would be of great assistance in modelling regions of less pronounced fluctuation in

traffic parameters such as speed and occupancy often observed prior to a shockwave actually propagating).

Although the parameters in the Gipps model correspond to characteristics of drivers and vehicles, calibration according to the situation modelled is required. This is because modelling of a motorway network and an urban network require quite different parameters for speed.

Three key assumptions were made in the formulation of Gipps' car-following algorithm (these also influenced his choice of initial parameters):

- a vehicle will not exceed its driver's desired speed at any time;
- free acceleration should increase with speed (as engine torque increases) then decrease to zero when approaching its desired speed;
- a driver will at all times maintain sufficient headway between his vehicle and the vehicle in front.

The equations derived by Gipps using these assumptions are reproduced in Appendix C. The equation for free-flowing traffic was obtained by driving an instrumented vehicle through traffic and fitting an appropriate relationship to the speed values derived. A vehicle in free-flowing traffic would be influenced by three factors all relating to itself, (desired speed, previous speed and maximum acceleration).

The equation describing behaviour during congestion conditions makes use of an additional parameter dealing with 'reaction time'. The effect of adding this term is to cause vehicles to commence braking earlier and to reduce braking as they approach the point at which they must stop, and thus creep slowly up to this point. Gipps believed that the behaviour of the vehicles was determined by the relative magnitudes of this variable and the apparent reaction time. If the additional reaction time was equal in magnitude to half the apparent reaction time then a vehicle travelling at a safe speed and distance will be able to maintain a state of safety indefinitely (providing the willingness of the driver ahead to brake hard has not been underestimated).

In real traffic this is not always the case and therefore when the willingness of the driver ahead to brake has been underestimated, the additional reaction time may have to be increased to avoid vehicles colliding. Gipps suggested that a smooth transition was possible between the two basic traffic conditions (described by Equations 1 and 4 in his paper). However, this does not occur when the leading vehicle brakes harder than the following vehicle has anticipated, or leaves the lane, or when a new vehicle moves into the gap between two vehicles.

Abou-Rahme and White also looked at the effect of estimating the braking rate. When $\hat{b}$ is less than b_{n-1} the term containing the speed of the vehicle ahead contributes more to the new speed of the vehicle, and the disturbances to the following vehicles are damped. This means that the driver has *overcompensated* for the velocity of the vehicle ahead. When the braking of the driver ahead is underestimated, these disturbances could be amplified.

Gipps suggested that three factors control the behaviour of the traffic,

- the distribution of desired speeds
- the reaction time for drivers, τ
- the ratio of mean braking rate to driver estimate of mean braking rate $\frac{\bar{b}}{\hat{b}}$.

This investigation looked at the sensitivity of the Gipps Equations to these three basic factors.

4.2.2 Stability within the Gipps model

The transition from stable to unstable car-following behaviour is important when considering the origin of a shockwave. Stable car-following behaviour is such that any perturbation to the behaviour of the lead vehicle is smoothed out as the perturbation propagates upstream. In order to reproduce shockwaves and observe them spreading out as they propagate, the stability of the car-following model must be considered, since it is likely that the spreading out of the perturbation is caused by unstable traffic.

The general formulation of a car-following model is

$$a_n(t+\tau) = \alpha[v_{n-1}(t) - v_n(t)]$$

where τ is the reaction time and α is a sensitivity parameter. May (1990) considered the behaviour of the following-vehicle and how that would determine the stability of car-following behaviour. Stable behaviour occurs when the following-driver is very attentive and therefore has a short reaction time. There is no over-reaction in the response and large acceleration or deceleration only feature during emergencies. A short reaction time and a low-sensitivity response are the characteristics of stable behaviour. If the response time of the following-driver is slow (so τ is large) and that same response is an over-reaction, there will be a significantly large acceleration or deceleration. A long reaction time and a high-sensitivity response are the characteristics of unstable behaviour.

May suggested that stability can be thought of in terms of a parameter, C which is the product of the reaction-time and the sensitivity parameter,

$$C = \alpha\tau .$$

There are two kinds of stability that occur in traffic flow, local stability and asymptotic stability. Local stability involves the lead vehicle and the next following vehicle only, whilst asymptotic stability involves a large (infinite) number of following vehicles. The variation in behaviour, both locally and asymptotically, depends upon the parameter C. Therefore, as an extension of the Gipps model this parameter was calculated and the effect of varying the value on the car-following behaviour was investigated.

Gipps remarked that the estimate of the braking of the driver ahead affects whether the signals are amplified or damped. If b_{n-1} is underestimated then the disturbances are amplified and vice-versa. Therefore it is intuitive that the value which corresponds to the stability parameter in the Gipps model is $\hat{b}$.

4.2.3 Estimating braking for the vehicle ahead

In his paper, Gipps estimates the maximum value of the braking parameter as $-3.0m/s^2$. In practice the estimate of braking is too small and so drivers react late and brake harder as a consequence. Within this investigation, the estimated braking parameter varied from $-2.5m/s^2$ to $-3.0m/s^2$ for all vehicles. The displacement-time plot representing this is shown in Figure 18.

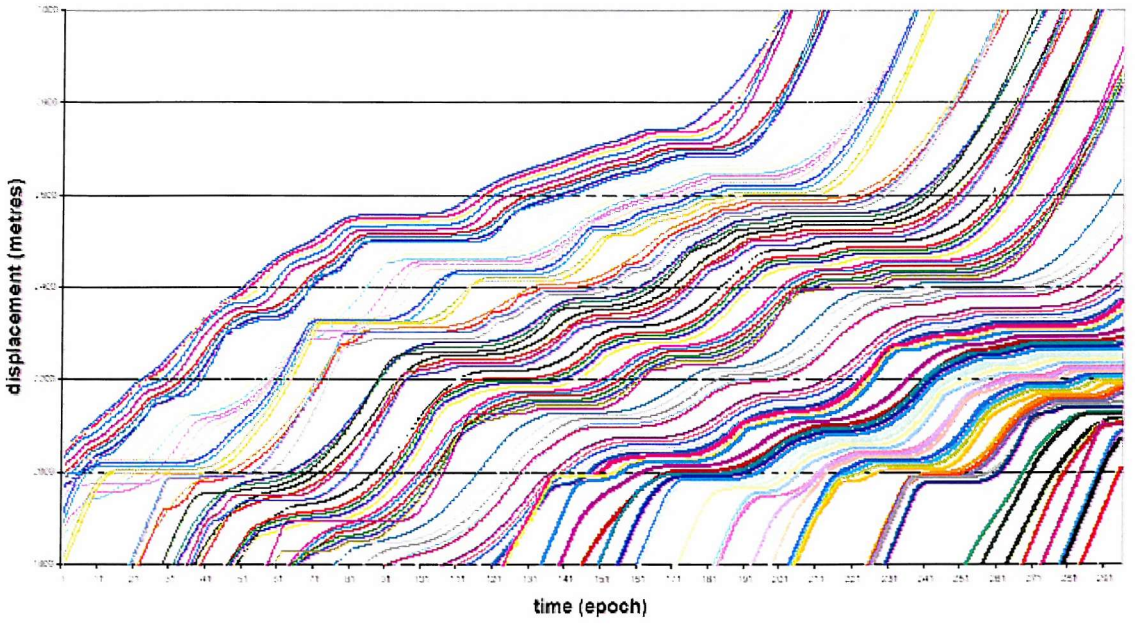


Figure 18 – Overlapping trajectories indicate collisions within the shockwave

Here the shockwave is spread out as it moves upstream and other shockwaves are created by some of the following-vehicles. A collision occurred because the final velocity of this vehicle was set too high and it moved further numerically than was physically possible! However, the problem of vehicles colliding had to be overcome. One method of overcoming this problem of collision was to increase the value of θ (the additional reaction time). The resulting displacement time plots show very smooth vehicle trajectories and no shockwave behaviour at all!

Alternatively, if the constraint is violated, the displacement of the vehicle can be altered so that the following vehicle does not collide with the vehicle in front. This represents a driver realising the need to brake harder and stop close to the vehicle ahead. As a result, the space that would normally have

been left for safety (according to assumptions made by Gipps) is actually encroached. Of course, in real life there will always be a chance that the driver does not brake hard enough and a collision does occur. Therefore if the inter-vehicular distance is too small the new displacement is calculated as:

$$x_n(t+\tau) = x_n(t) + x_{n-1}(t+\tau) - x_{n-1}(t)$$

when the velocity of the vehicle is above 5m/s. This means that the vehicle is allowed to move as far as the vehicle ahead has just moved, but no more. When the velocity of the vehicle is below 5m/s the vehicles are allowed to move bumper to bumper and so the new distance is:

$$x_n(t+\tau) = x_{n-1}(t) - s_{n-1} + \eta$$

where η (in metres) is a constant between 1 and 2, representing the extra length added to the vehicle length to make the *effective* vehicle length. By introducing this value the assumption (A3) from Gipps is relaxed. Recall that Gipps stated that the extra safety reaction time was required in order that vehicles stopped slowly behind the vehicle ahead and did not breach the extra space in the vehicle length. Observed behaviour does not always support this assumption, since drivers often come too close to the vehicle ahead and need to correct their speeds.

The modification now included in the model means that in high density flow the braking behaviour reproduced is more true to life. The new velocity is calculated as

$$v_n(t+\tau) = \frac{2}{\tau}\left(x_n(t+\tau) - x_n(t)\right) - v_n(t).$$

The results obtained using the new displacement and velocity calculations are presented in Figure 19. The two shockwaves on the right of the picture are propagating from the disturbance to the profile of the lead vehicle. The shockwave on the left has not been caused by the lead vehicle, but by disturbances being amplified within the traffic flow. Slight perturbations in the trajectory of one vehicle lead to larger ones reproduced for the following vehicles. A greater error in estimated braking will lead to a greater amplification of these perturbations.

New shockwaves were also created in the trajectories of the following vehicles, not directly related to the behaviour of the lead vehicle. It is this type of seed-point which is the basis of the Mechanistic Approach.

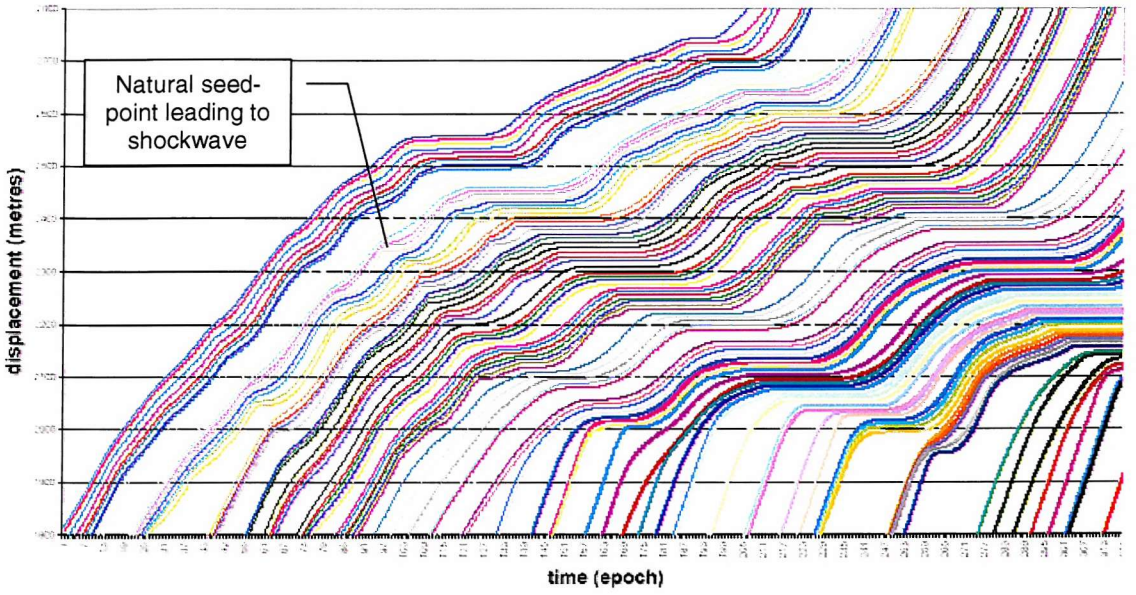


Figure 19 – Example of shockwave generation using Gipps Algorithm

Focussing on the question of effective amplification, the work conducted by Abou-Rahme and White led to extensions in the modelling equations to incorporate the following additional features:

- inertia for vehicles that have been stationary for some time and take longer than normal to respond to the vehicle ahead moving off;
- anticipation of brake lights ahead when approaching a slow-moving section, including an awareness parameter for every driver.

Incorporating additional equations to model anticipation of brake lights had significant effect on the shockwave behaviour. The shockwave is seen to spread out as it moves upstream and only dissipates when there is insufficient demand, that is, when vehicles are too far apart to be influenced by the behaviour of the vehicle in front.

4.2.4 Modelling seed-points

The results from both the simple model and the microscopic model showed occasional shockwaves occurring independently of those which had been induced for the purpose of demonstrating car-following behaviour. A closer investigation of these was conducted with a view to understanding what properties within the car-following model would replicate seed-points.

Initial analysis showed the independent shockwave within the microscopic model to have been caused by a collision, and it was a curious feature of the model that changes to the car-following algorithm had in fact increased the overall number of collisions. This was because of necessary changes to acceptable headway settings and braking estimations allowing a vehicle to come closer to the next vehicle than would be comfortable and then brake harder (creating a more realistic oscillating motion). Further investigation of how the microscopic model dealt with vehicle braking showed a lack of distinction between 'maximum desirable braking' and 'maximum possible braking' rates. The braking rates used within the model must also be updated and this would precipitate a re-calibration of the whole model, activities that are outside the scope of this thesis.

Improvements to how driver perception is represented in the microscopic model led to the identification of another important deficiency, namely that the driver assumes the braking rate of the vehicle in front is identical to his own! This unrealistic assumption does in fact counter-balance the number of collisions overall, but it is not a satisfactory assumption. It is recommended that the revised car-following algorithm (which assumes the braking rate of the vehicle in front is related to the class of the vehicle in front) be implemented and alternative solutions to the high rate of collisions be investigated instead.

The purpose of this exercise was not to promote the merits of Gipps, but to demonstrate that seed-points might occur independently within a car-following model. Although Gipps claimed that the equations would never allow vehicles to get too close, a constraint was included to affirm this and necessary extensions made by Abou-Rahme and White to overcome situations where the constraint was violated. Propagation of shockwaves was demonstrated in

motorway traffic and the model equations were modified to improve replication of this behaviour. Relaxing the assumption that vehicles will never get too close has also improved the reproduction of shockwaves and even caused some shockwaves to be generated.

In conclusion, it was possible to replicate the origin of a shockwave within a car-following model by varying the stability parameters. This result provides an important input to the modelling of seed-points, but forms only part of the overall picture. Although this particular car-following model could accommodate shockwave origins along with shockwave propagation, much depends on the environment in which the car-following algorithm is used (for example, how it relates to a parallel lane-changing algorithm). If a full implementation of the car-following algorithm becomes available within a microscopic model (such as SISTM) then replication of seed-points using this method can be tested in detail.

Brackstone and McDonald (2003) considered four pertinent issues relating to the successful modelling of traffic flow and driver behaviour, including the relevance of safe following-distances, the use of time increments in microsimulation, and the role of asymptotic behaviour in generating flow-breakdown within a model. Clearly the challenges expressed in that paper need to be addressed if progress is to be made in modelling the origins of flow-breakdown using car-following theory.

4.2.5 A deterministic extension

Observation of traffic data on the M25 has shown that a seed-point also has a distinct 'location-based' property, in other words shockwaves frequently arise from a specific location. Whilst the Mechanistic Approach is useful in modelling the actual generation of the shockwave, it does not allow any kind of control with regard to where the shockwaves occur.

In Chapter 3, a number of possible causes of congestion were identified based on an empirical study for the M25. The mechanism common to each of

these cases is a *redistribution* of headway within a cluster of proximate vehicles. For example, as vehicles negotiate a curve in the alignment some drivers will tend to accelerate into the bend, some will cruise at constant speed and some may lose speed. The headway will change as a result of passing this location and for *no other reason*. In order to model such behaviour, a number of properties must be considered such as driver aggressiveness, vehicle power and so on.

A particular *location* could therefore be classed a seed-point. The effect of such a location on vehicles passing it should be to force a redistribution of vehicle spacing according to the logic described above.

Once redistribution has occurred, a certain number of vehicles may find that they are now too close to the vehicle ahead and thus brake to compensate. If the appropriate traffic conditions exist then this will lead to the propagation of the shockwave from just downstream of that location. The first indication of any sustainable speed drop will be a few hundred metres upstream, namely at the geometric feature itself.

Two points are worth noting at this stage. Firstly, a significant amount of calibration would be required in order to extract parameters that led to a satisfactory number of disturbances being created at the specified seed-point. This could only really be tested and validated once a satisfactory microscopic model was in place and replicating shockwave propagation. Secondly, the approach is only useful if the *focus* is on shockwave modelling, since the predetermination of a location as 'shockwave prone' may be restrictive to the versatility and performance of a microscopic model in the long term.

However, neither of these considerations prohibits the deterministic approach from being taken forward.

4.3 THE PROBABILISTIC APPROACH

4.3.1 Inferential and Bayesian Statistics

When dealing with headway distribution for a platoon of vehicles, it is common to think in terms of probabilities. Luttinen (1994) presented a four-stage

approach to determining the best-fit shape of headway distribution, depending on traffic conditions. The paper also highlighted some of the problems typical of research in this field. The exponential distribution was the preferred model for extremely low flow conditions. The gamma distributions could be used if simplicity is a requirement and if there is empirical support for short headways having a low probability. The log-normal distribution falls between the simple non-realistic and complex realistic models, but it has theoretical advantages as a follower headway distribution. The generalised Poisson distribution was recommended for demanding applications with adequate computational facilities, in particular for looking at vehicle arrival rates.

Inferential statistics provides a mechanism for incorporating new information (from empirical observations) with prior information about random events. The link between the prior information and the empirical data is provided by likelihoods that are also probabilities. The result is a posterior distribution that **reflects both the empirical data and the prior information**.

Thomas Bayes (whose work was published posthumously in the Philosophical Transactions of the Royal Society of London in 1764) first documented this principle in 1761. The introduction stated that the purpose of his approach was:

> "...to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times".

The classical approach to inference is the estimation of an unknown parameter (θ) by that value which makes the data set most likely, in other words the maximum likelihood. The overall framework within which Bayesian inference works is conceptually identical to that of statistical inference. There is a population parameter θ about which an inference can be made, and a probability mechanism $f(x \mid \theta)$ which determines the probability of observing different data x, under different parameter values for θ.

The fundamental difference to classical inference is that θ is treated as a random quantity. Inference will be based on the probability distribution of the parameter given the data $f(\theta \mid x)$ rather than that of the data given the parameter $f(x \mid \theta)$. This leads to more natural inferences, but to achieve this it is necessary to specify a prior probability distribution $f(\theta)$, which represents a certain belief about the distribution of θ prior to having any information about the data. This notion of a prior distribution for the parameter θ is at the heart of Bayesian thinking.

In its most basic form, the Bayes Theorem is a simple result concerning conditional probabilities. If A and B are two events with Pr (A) > 0 then

$$\Pr(B \mid A) = \frac{\Pr(B)}{\Pr(A)} \times \Pr(A \mid B)$$

This relationship is expanded in terms of probability density functions (pdf) allowing an inference to be made about the unknown parameter on the basis of prior knowledge and the data. The application of this study will therefore be to make a prior estimate of the flow-breakdown parameter (θ), and then use the observed data in order to improve the estimate.

4.3.2 Applying the model to traffic

In Section 2.4 the conceptual basis for this thesis was attributed in part to the work of Elefteriadou, Roess and McShane (1995). The paper gave a good critique of conventional attempts to identify flow-breakdown using the speed flow curve. It also presented the problem of flow-breakdown at merge points in terms of a chain of events with assignable probabilities that could be summated. Elefteriadou (1997) refined the process for identifying one of the key events in that sequence (the arrival of vehicle clusters) by using Poisson distribution. This is a common application, assuming that the time between vehicles (or headway) is exponentially distributed such that the probability that there is an event during time T is $(1-e^{-\lambda T})$. This is acceptable for periodic arrivals (such as ramp flows or light mainline flows), but in the specific case of heavy motorway flow an improvement on this approximation is required. For example, the Poisson process assumes independent arrivals, but the concept

that traffic flow may consist of two classes of vehicles, (constrained and free-moving) was proposed as early as 1955 (see May 1990 for a review). Haight, Whisler and Mosher (1961) examined the methods used to describe the distribution of cars on the road, showing the difference between headway during free-flowing traffic and headway in the lead up to congestion. Buckley (1968) proposed that the only inhibition to the underlying Poisson traffic process is the existence of a so-called 'zone of emptiness'. If this zone is distributed in time as g(t) and if the headways greater than the random variety sampled by g(t) are distributed exponentially as h(t), the composite pdf for headway becomes:

$$f(t) = \phi\, g(t) + (1-\phi)\, h(t) \quad \text{where } 0 < \phi < 1$$

The equation is dominated by g(t) for small headway, and by h(t) for larger headway. Banks (1995) also commented on this distinction in his review of the assumed relationships among traffic flow characteristics, as he developed a dual-density model for the traffic data.

The natural implication of the above discussion would be to start seeking an improved vehicle arrival model. Figure 20 shows sample one-minute averaged flows for a particular location along the M25 Controlled Motorway (with height indicating flow value for that minute, and colour indicating the corresponding speed). Figure 21 shows a similar data set but for a different day. At first glance the profiles appear to be the same, but closer examination shows the daily variation to be significant. Although it would be generally useful to have a statistical model for the probability of finding a particular flow at a specified time, this does not address the problem in hand.

A probabilistic model for the origin of motorway shockwaves requires a detailed description of the conditions present at the **start** of flow-breakdown in order to make an inference about the statistical relationship involved. By

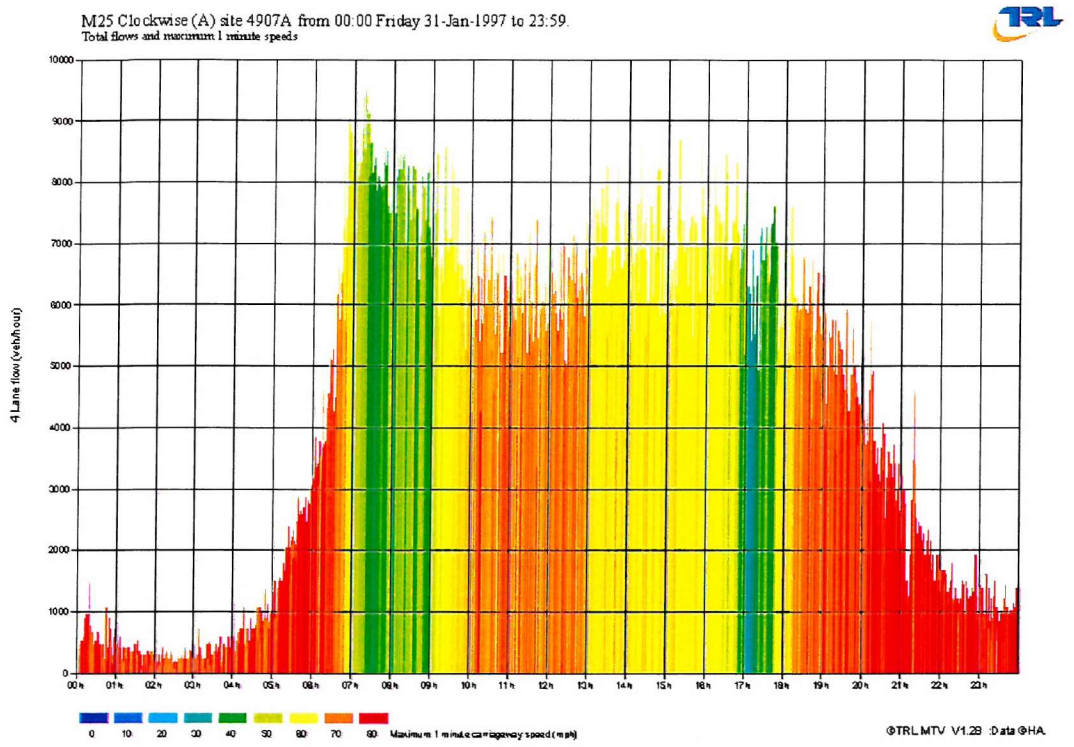


Figure 20 – Sample one-minute flows from M25 for 31 January 1997 at detector 4907A, displayed using MTV

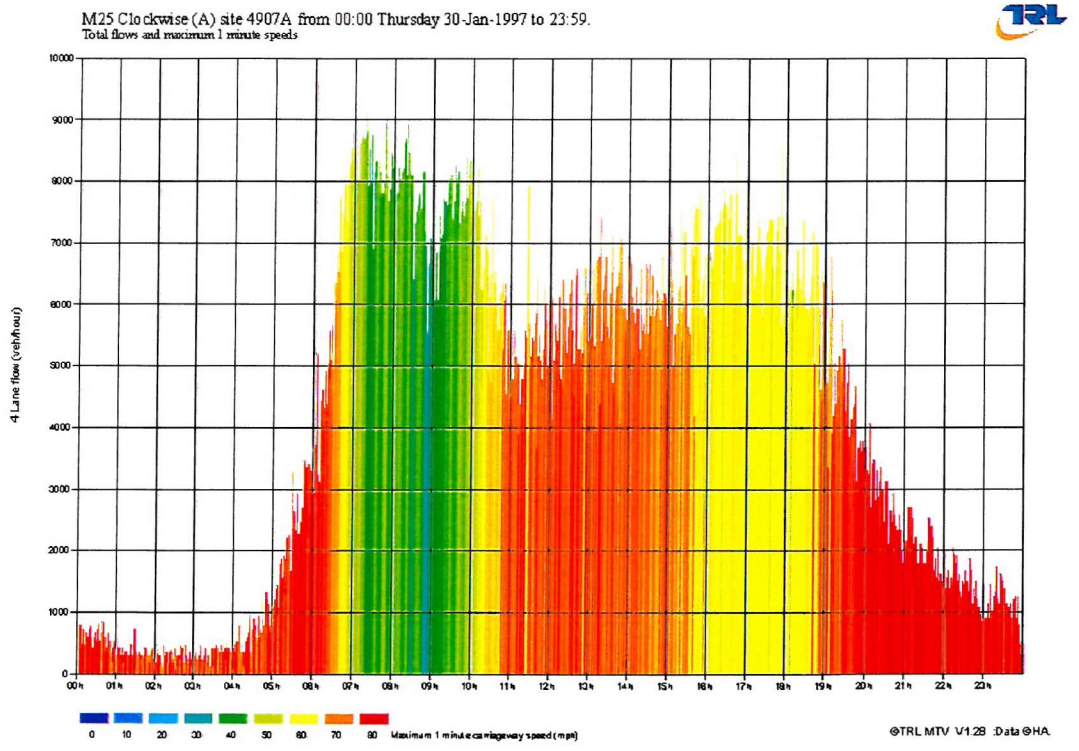


Figure 21 – Sample one-minute flows from M25 for 30 January 1997 at detector 4907A, displayed using MTV

denoting the flow-breakdown variable as (θ), and the specific traffic parameter as (q), then in Bayesian notation the model becomes:

$$f(\theta \mid q) = \frac{f(q \mid \theta)\ \pi(\theta)}{\int f(q \mid \theta)\ \pi(\theta)\ d\theta}$$

Once it has been integrated, the denominator for the right hand side is not dependent on θ and therefore could simply be written as *f*(q). The prior pdf π(θ) is a statement of belief about the occurrence of flow-breakdown. By integrating both sides with respect to θ during the interval (0, 1) it becomes possible to evaluate the denominator.

The difference between this work and that presented by Elefteriadou is therefore in the definition of probabilistic events. The focus is on traffic states and how these are classified, rather than on vehicle arrival times and flow rates. The next section expands on how the flow-breakdown variable θ is treated.

4.3.3 Specifying the prior function

The prior function in this context is a statement of our present beliefs on flow-breakdown (θ). Classification of flow-breakdown is a subject of much debate amongst transport professionals and there is no straightforward answer. In the next chapter this is given further consideration and a systematic approach to classifying seed-points is presented.

Kerner (1999) offered a helpful discussion around this subject and referred to an intermediate state between free flow and congestion called 'synchronised flow'. He also proposed further subdivisions in terms of stability. The existence of an intermediary state is implicit within this current study and fundamental to the proposed model. The flow-breakdown variable (q) is used to distinguish between this **intermediate state** (the beginning of congestion) and the remaining two states (free-flow and congestion) both of which are irrelevant to seed-points. However, the boundaries between these traffic states are also defined by the detection method adopted. The prior function is

based upon **a binary variable that operates within the constraints of a predefined search space**. See Chapter 5 for further detail.

The prior function can be described as a Bernoulli distribution with respect to the variable (θ) as follows:

$$\pi(\theta=1) = (p)$$
$$\pi(\theta=0) = (1-p)$$

For a sample size of (n) possible points containing (r) successful outcomes and (n–r) unsuccessful outcomes, the probability of (r) outcomes can be expressed as:

$$p^r[1-p]^{n-r}$$

There is a finite and definable number of combinations for such a sequence of events to take place. Using the common notation the equation becomes:

$$^nC_r\ p^r(1-p)^{n-r}$$

or the Binomial Distribution. In this case, the estimate of $\pi(\theta=1)$ simply becomes $(\frac{r}{n})$. With inferential statistics, the outcome does not depend entirely on the choice of prior function, and dependency is inversely proportional to sample size. The assumptions taken here are reasonable, and can be improved by the size of the data archive being examined.

4.3.4 Selecting a likelihood function

The likelihood function will be derived from the database of seed-points. It is a description of the distribution of traffic variables associated with the seed-point (according to the definition in Chapter 5). Corresponding values associated with non seed-points are also required. Because of the data resolution it should be possible to work with the **discrete values** for frequency distribution without resorting to standard functions. However, there may be circumstances where it is expedient to do so, therefore a brief discussion of suitable functions is required here.

For traffic analysis the generalised Poisson distribution is usually considered appropriate for looking at random arrivals during low flow conditions or the distribution of a single traffic parameter about a critical point. But the seed-point detection process gives rise to a slightly different type of distribution.

Early examination of the data sets obtained through this process showed signs of a possible normal distribution for the individual parameters. The distribution of speeds associated with seed-points followed a bell-shape, and the distribution of occupancy followed a similar pattern.

In practice the probability of a particular pair of values occurring would simply be the product of the two likelihoods obtained through the normal distribution of the independent variable. But there are two further points to consider:

- the implicit relationship between speed and occupancy (in this case a proxy for headway);
- the dependency on both variables by the algorithm developed for seed detection (how the pair are grouped in the first place).

The distribution of variables associated with the seed-point must therefore be examined as a coordinate pair (rather than as independent variables) and the extent of that dependency must be taken into account for the remaining calculations.

As both variables displayed an apparent similarity to the normal distribution it was logical to consider the suitability of the Bivariate Normal. This distribution is expressed as follows:

$$N(k, v) \sim \frac{1}{2\pi\sigma_k\sigma_v\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left(\frac{1}{1-\rho^2}\right)\left(\frac{(k-\mu_k)^2}{\sigma_k^2} - 2\rho\frac{(k-\mu_k)(v-\mu_v)}{\sigma_k\sigma_v} + \frac{(v-\mu_v)^2}{\sigma_v^2}\right)}$$

which is distinguished by the presence of an additional variable ρ (the coefficient of correlation). The value of ρ is zero when the covariance between k and v is zero, making the above equation equal to the product of N(k) and N(v).

Figure 22 gives an illustration of the bivariate normal distribution.

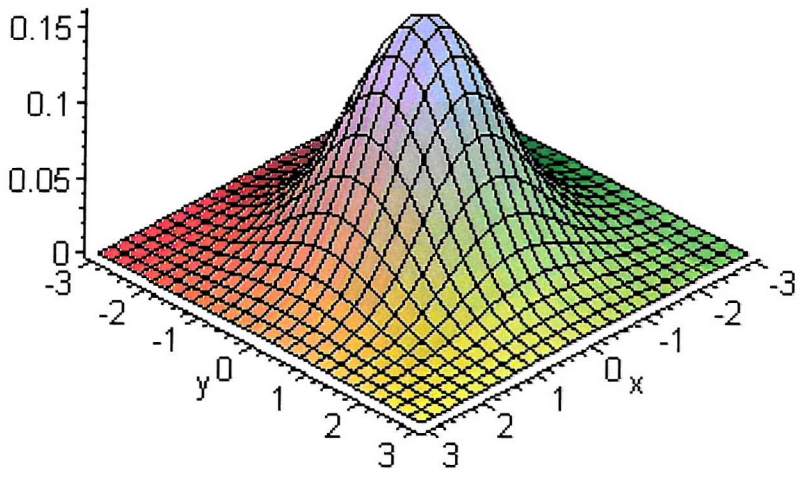


Figure 22 – An illustration of the double-bell bivariate normal distribution

Five statistical parameters therefore need to be calculated – namely the means, standard deviations, and covariance for speed and occupancy (μ_k, μ_v, σ_k, σ_v, and ρ_{kv}).

4.3.5 Summary of the Bayesian Model

The Bayesian Model allows us to derive a posterior distribution for the probability of flow-breakdown occurring given that a traffic data (q) has been detected. Since θ is in fact a binary and discrete variable, it is possible to rewrite the equation solving the integral as follows:

$$f(\theta=1|q)=\frac{f(q|\theta=1)\cdot\pi(\theta=1)}{[f(q|\theta=1)\cdot\pi(\theta=1)]+[f(q|\theta=0)\cdot\pi(\theta=0)]}$$

the denominator being f(q) a distribution of the traffic variables throughout the data set independent of θ. Finally, it is evident that the sum of the two probabilities describing the occurrence of flow-breakdown equals 1 so a simple substitution gives:

$$f(\theta=1|q)=\frac{f(q|\theta=1)\cdot\pi(\theta=1)}{[f(q|\theta=1)\cdot\pi(\theta=1)]+[f(q|\theta=0)\cdot(1-\pi(\theta=1))]}$$

The prior function in this case is taken to be a simple ratio. This value is applied through the equation to every 'combination' described within f(q | θ=1). Effectively the prior function should then be written as a matrix (with the same value in every cell), and applied to every combination of (v, k) within the search space. Once the posterior is obtained it is possible to exploit the key feature of Bayesian Inference, using **this posterior as a refined prior** for a new set of (similar) data.

4.4 SUMMARY FOR MODELLING

This chapter has proposed two approaches to modelling seed-points and has explored the feasibility of both. Investigation within the Mechanistic Approach was limited mainly by the availability of a microsimulation model that could adequately represent shockwaves. It was hoped that implementation of the refined car-following model proposed by Abou-Rahme and White (1999) within SISTM might have yielded such a tool, but this has been delayed by a series of external factors.

The rest of the study will now concentrate on implementing the Probabilistic Approach. The basic form of the statistical model has already been presented, and this will be used to analyse the data set obtained through an incident-detection and classification approach described in the following chapter.

5 DETECTING SEED POINTS

5.1 THE NEED TO DETECT SEED POINTS

The previous chapter described the development of a probabilistic model using a Bayesian approach to accommodate prior belief about shockwave occurrence. In terms of **data requirement**, the probabilistic model uses a detailed description of the current conditions during which flow-breakdown occurs, in order to make a statistical inference about the relationship involved.

Detection of possible seed-points in the traffic data is necessary to provide an estimate of the prior function (describing the frequency of flow-breakdown) and to provide a distribution of values for traffic parameters associated with that flow-breakdown. The approach to detection must therefore be able to distinguish between the origin of a shockwave and the main body of a shockwave, and then record the associated traffic data (in particular the one-minute speed, flow and occupancy values).

To address this requirement and facilitate numerical analysis, a **software utility** entitled SeeDetect© was developed by the author using Visual Basic. SeeDetect© reads traffic data from binary files and stores it in predefined arrays, separating the information by lane and calculating the lane average as an extra part of the array. The software core comprises a series of modules for testing incident-detection algorithms and a special module that holds the main algorithm for seed-point detection. Notable features include the ability to visualise shockwave patterns, to identify seed locations (either automatically or manually), and to store comprehensive information about the seed-point in various database formats.

Figure 23 shows an actual screenshot derived from the current version of SeeDetect©, with a selection of seed-points marked as coloured rectangles. The software has been set up to load MIDAS traffic data (speed, flow, occupancy) in blocks of six-hours (the day can then be covered in four viewings). There are simple controls for scrolling back and forth across successive six-hour blocks within the same day, and the relevant time interval

is displayed on the screen. The 'manual select' button provides the ability to select the seed-points visually using the mouse (useful as part of the initial learning phase). The software is also capable of displaying information about each cell and marking the seed-point and its surrounding data region.

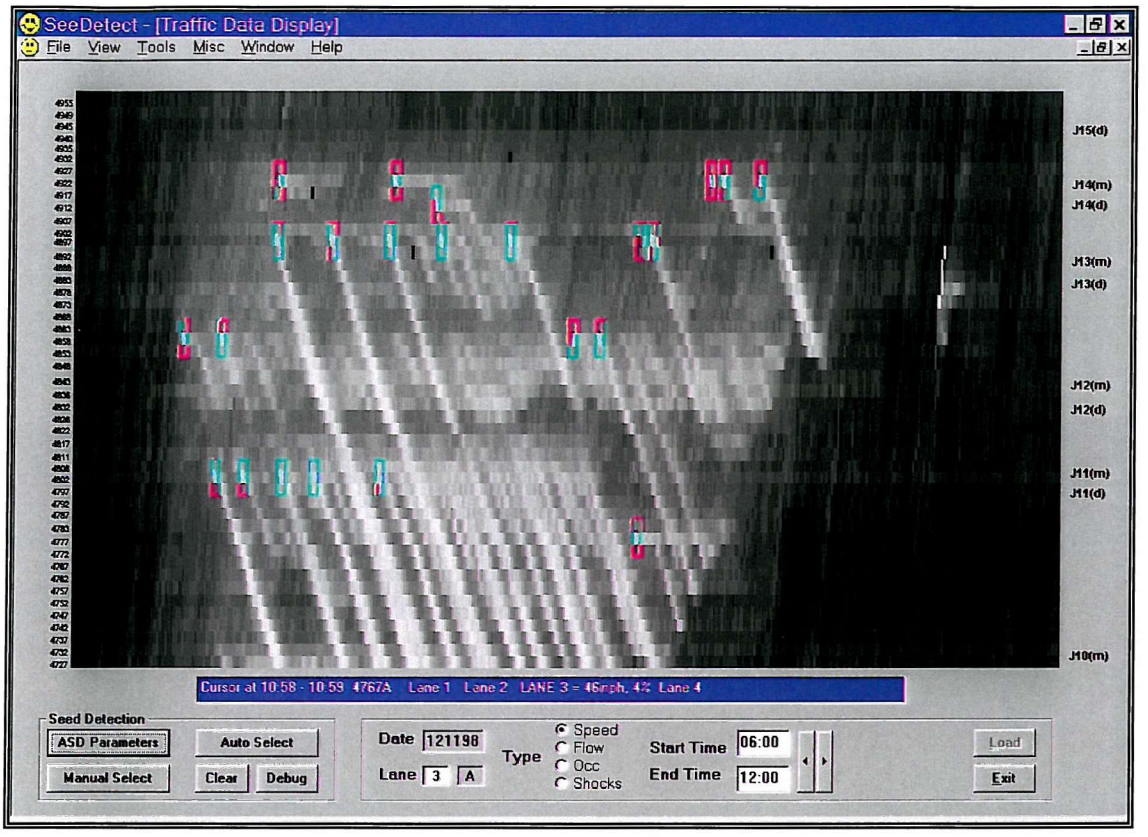


Figure 23 – Some possible seed-points identified using 'Manual Select' feature, against a greyscale speed background, (shockwaves appear as white streaks)

For each selection, a full catalogue of information is sent along with an index marker to a database, and this information can be accessed directly or through SeeDetect©. The output currently includes the date, time, position, carriageway, lane number, and speed/flow/occupancy at the seed-point and the surrounding locations in the data map.

As well as processing the seed-points, the software allows the user to examine the shockwave tails. For every complete shockwave object (a seed and tail combination) additional properties for the shockwave are also stored (for example the speed of propagation). All this information stored in a database, and can be reloaded onto the diagram at a later date if required.

5.2 INCIDENT DETECTION ALGORITHMS

As part of this study a special module within the software was developed by the author to use incident-detection algorithms for filtering traffic data to find shockwave patterns. The idea was that seed-points could then be located from the shockwave body. A selection of algorithms was reviewed for suitability.

Figure 24 gives an example of the output derived through the California #8 Algorithm. California Algorithms differentiated well between incidents and shockwaves although they generated a fair amount of noise once parameters had been optimised for shockwave detection.

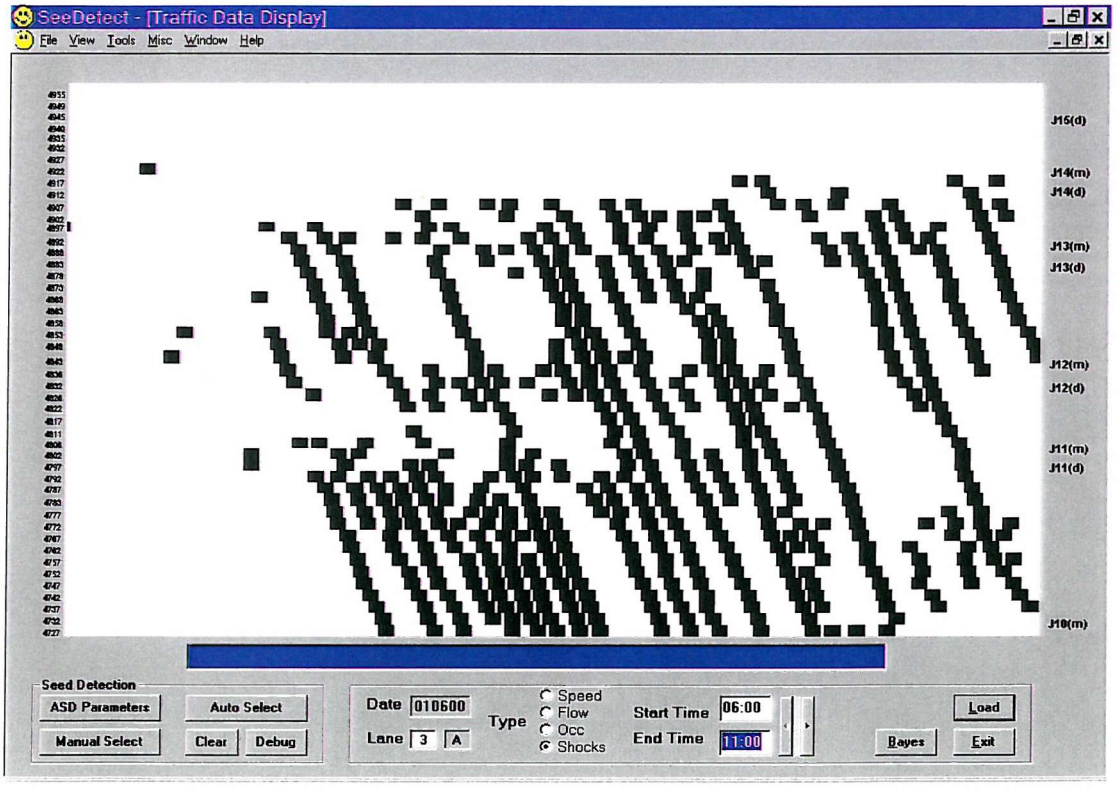


Figure 24 – Shockwaves identified by using the California Algorithm filter

The McMaster Algorithm provided a more consistent detection of shockwaves with little additional noise, but it was harder to implement than the rest of the algorithms because of the requirement to calibrate against speed data. The exponential family was discarded because the outputs were unrepresentative of the traffic patterns except occasionally at the shoulders of the peak period.

Of the other algorithms reviewed, one could not identify the shockwave until well upstream of the seed-point, and the rest could not detect the shockwave consistently.

The California #8 was taken forward for first use in the traffic data pre-processing phase. The algorithm was applied to one-minute data in order to improve response time; however, some degree of averaging or smoothing was required to overcome the erratic elements found within raw data. By definition there was a time lag due to the time-series calculations, and this introduced a degree of approximation into the method which was not satisfactory. The top of the detected shockwave still fell short of the seed-point by one (or sometimes two) detectors, this point being the result of two or three minutes of propagation. A more suitable method therefore had to be developed.

5.3 TOWARDS AUTOMATIC SEED POINT DETECTION

One of the main arguments early in this thesis was that a seed-point is an event triggered by a combination of traffic characteristics *and* location effects, the two components are necessary for occurrence, (see Section 1.1). The empirical study in Chapter 3 considered seed-points in terms of location and examined some of the causal factors behind shockwave occurrence. In considering the question of automatic seed-point detection using classification of traffic parameters to describe certain characteristics, it becomes necessary to treat the seed-point as **an event with a particular signature** in the traffic data. There is a change of emphasis in the remainder of this thesis from the *location* aspect to the *event* aspect, and data analysis is conducted independently of location.

The next step was to use a classification approach in order to define a filter for seed-points using the available traffic data. An early part of this process involved scanning for speed differentials across two consecutive loop detectors exceeding a pre-defined threshold. The process was refined through introducing a filter for minimum flow and occupancy in order to

manually define the 'upstream conditions' for which it was anticipated flow-breakdown would take place.

Based on observations within the traffic data, assumptions were made about the speed of shockwave propagation, as this would help plan which cells should be checked. A set of rules was developed to examine the traffic-data array and identify discontinuities which might lead to shockwaves.

Figure 25 is a schematic representation of a shockwave commencing at time t+1 at detector 4907A and propagating upstream through 4987A and beyond as time increases. Each 'cell' shown here would contain a value for speed or flow during that minute at that detector (compare with Figure 23). The gradient represents the speed of the shockwave moving upstream (passing through 1km of traffic in just over 3 minutes).

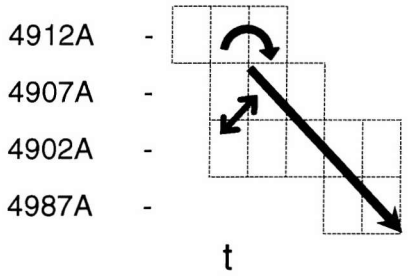


Figure 25 – A schematic representation of a shockwave moving through the position-time array, and the information attainable at various cells.

The double arrow shows the cells used to identify the speed-drop (assuming vehicles could travel at free-flow speeds when passing 4902A). Every entry in the position-time array was then placed into one of four categories (flow-breakdown, minor fluctuations, downstream seed-point, and no speed drop), and the intention was to track the shockwave origin using this output.

Significant time was devoted to developing these rules but the early output only provided a partial identification of shockwaves (along with some incorrect results). The exercise did ,however, bring a valuable understanding to the task of defining a seed-point. Looking again at the seed-point in context, the

following convention was used to store the traffic information (the '0' in this diagram indicates the position of the mouse click using SeeDetect©):

8	1	2
7	0	3
6	5	4

The relationship between information in Cells 5–7 and Cell 0 was of possible interest for application of the probabilistic model.

One of the most important features was the fact that the selected seed-points seemed to occur at the threshold where speed was falling and traffic density was rising. Figure 26 provides a schematic illustration of this logic.

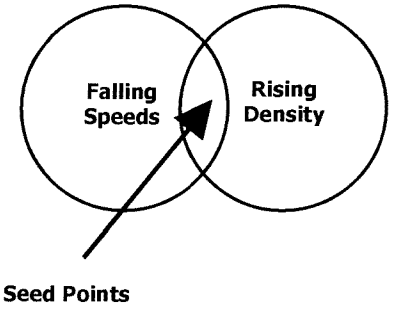


Figure 26 – The overlap of rising density and falling speed

Figure 27 compares the clarity of shockwaves within traffic data, contrasting flow against speed and occupancy. Much of the subtle variation associated with the regions from which shockwaves originate is missing in the flow plot. In terms of information about the location of a seed-point, the flow variable is comparable to the output from the incident detection algorithms back in Figure 24. As there was no evident link between traffic flow and the occurrence of seed-points, the study **focussed on speed and traffic density** as the identification variables.

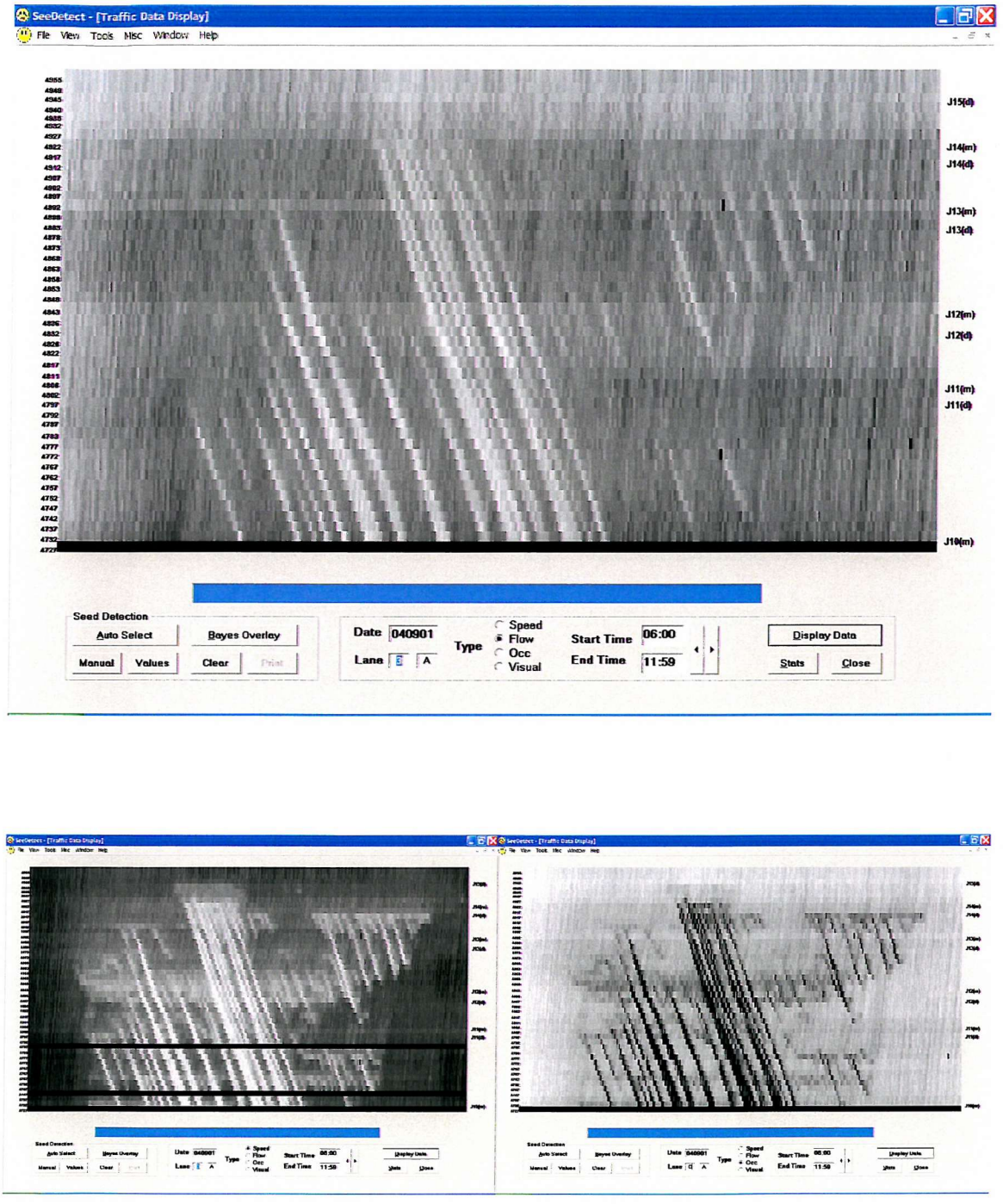


Figure 27 – Clarity of shockwaves within traffic data – the top plot shows flow, the two smaller plots show speed and occupancy respectively. Much of the subtle variation associated with the origin of the shockwaves is missing in the top plot (flow). This was consistent throughout the data examined.

Profiles for speed and occupancy at specific loop sites were examined and compared against MTV Plots showing the shockwave in detail. This suggested that one of the characteristics of a seed-point is the numerical proximity of the speed and occupancy values (for example, see Figure 28). Although this is a coincidental finding (based on the units selected in this instance) it does offer some assistance with placing coordinate pairs within the overlapping region shown in Figure 26.

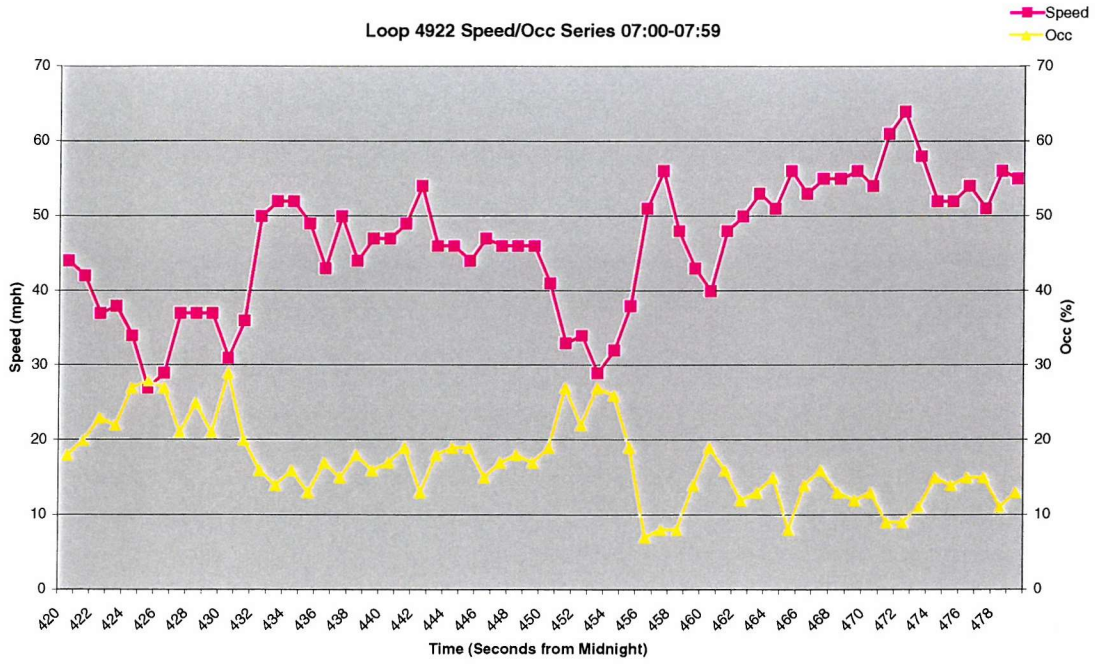


Figure 28 – Comparing values of speed and occupancy at M25 Loop 4848

Application of this understanding led to the following rules being implemented as part of the processes of detection within the software (see Appendix F for code listing)

- verification that the occupancy value lies within a predefined range
- verification that the speed value also lies within a predefined range
- time series comparison of speed and occupancy, to verify a rise in occupancy and a simultaneous fall in speeds
- additional check on numerical proximity of speed and occupancy using these particular units, to assist with classification
- final check for the direction of the gradient on speed and occupancy profiles, (to distinguish the seed-point from the edge of a shockwave).

The rules are a function of speed and occupancy data readings within the 3x3 operand described above. It is important to note that there is also an implicit assumption about the existence of 'synchronised flow'. This is described by Kerner (1999), who recognised that the transition from 'free flow' to 'congestion' is not abrupt but via a third (and overlapping) phase. The characteristics of traffic at seed-point locations selected during the rule development phase (described above) are comparable to the synchronised flow described by Kerner (falling speeds and rising density of traffic).

Since the purpose of detection within this study is to collect data samples with confidence the above rules were implemented as an algorithm and taken forward. Effectively these rules would define a seed-point for the purpose of this study. The database would therefore contain records of seed-points (defined as locations which satisfied all five rules) and a corresponding negative set (locations with similar values for traffic, but not qualifying as seed-points).

However, the algorithm also required thresholds in order to operate. Some initial estimates were taken from the data already gathered through manual selection, (and compared with MIDAS data as displayed in Figure 28 above). But further validation was required before the rules could be used.

5.4 SPEED-HEADWAY RELATIONSHIPS

5.4.1 Introductory comments

In the previous section it was found that **two variables were essential** to the seed-detection algorithm, namely average speed and loop occupancy (both calculated over a one-minute period). The occupancy variable is essentially a proxy for vehicle headway, and it was possible to explore the speed-headway relationships in real traffic data from the M25.

The work by Helbing (1997) fitting a theoretical function to individual vehicle data collected on a Dutch two-lane autobahn has been considered. Helbing took the standard continuity equations and incorporated an equilibrium velocity term. The resulting relationship included speed terms on both sides

of the equation and can therefore only be solved numerically. Helbing himself concluded the model is only useful for curve reconstruction at high densities.

The speed-headway function forms an integral part of a large number of car-following algorithms. Holland (1998) provided an excellent summary of the composition of models including Edie, Chandler, Herman and more recently that proposed by Bando. Although Holland focussed on developing a general stability criterion, he provided a table of the "effective speed-headway function" implicit within the models he reviewed.

5.4.2 Collection and analysis of individual vehicle data

This section describes how vehicle data was acquired from key locations on the M25 and how it was analysed in order to derive a basis for the rules used in SeeDetect©. A data collection exercise was conducted through projects led by the author at TRL Limited and analysed for this study. The availability of one-minute averaged traffic data from the MIDAS System on the M25 has already been described. However, by connecting directly to the MIDAS Outstations using laptops with special software, it was possible to obtain an additional level of information known as Individual Vehicle Data (IVD). A total of three weeks of data was collected on M25 between J13 and J16 within a period of three months (April to June 2000). The data set includes periods of the day when the variable speed-limits were in operation.

For each of the speed groupings between 0 and 80mph {0–1mph, 1–2mph, ... 79–80mph}, the number of vehicles with a time headway in each of the groupings {0–1sec, 1–2 sec, ... 49–50 sec}, was calculated. Upper limits of 80mph and 50 seconds were chosen arbitrarily to simplify the data analysis, as values beyond this range do not affect the present analysis. For each of the speeds up to 80mph the mean and mode time-headway (for every value of headway, not just those under 50 seconds) were calculated.

All of the above was done on a lane by lane basis, and for all lanes aggregated together. For a given speed, the average distance-headway was calculated by multiplying the average time-headway by that speed. The midpoint of the speed grouping was used (for example, a value of 21.5mph

was used for grouping 21–22mph). Figure 29 is a plot of speed against time-headway calculated for the nine days with cleanest data, across all four lanes at loop 4902A (M25 J13–14). Data was obtained for the full 24 hours on each day hence the scatter in headway values at the higher speeds.

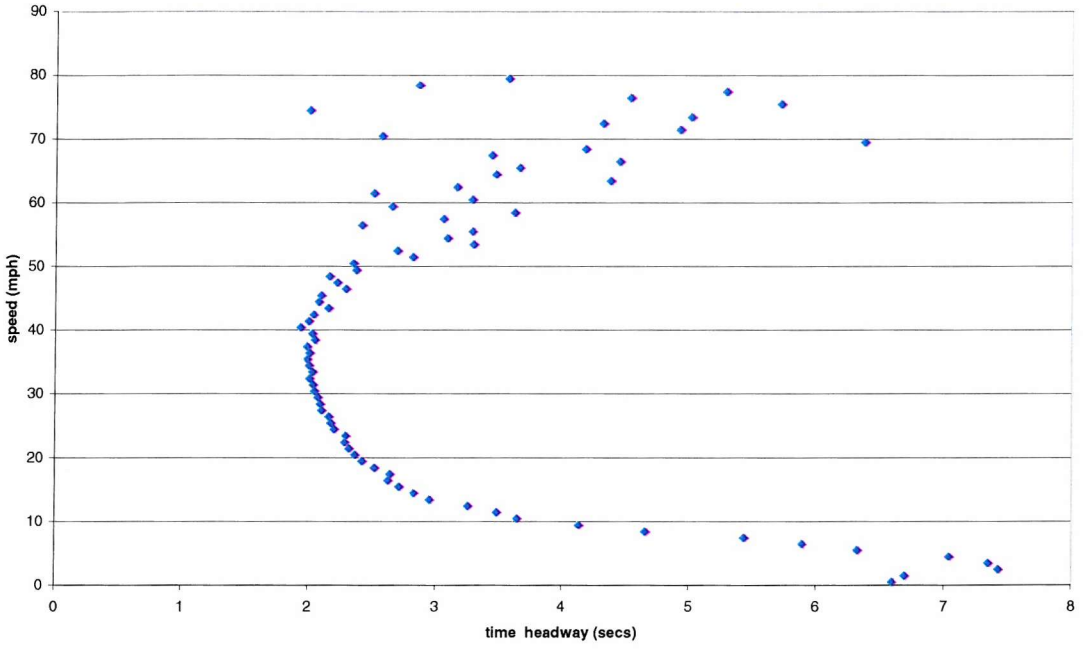


Figure 29 – Speed against time-headway for 9 days at Loop 4902A (average of all lanes)

The critical time-headway (using data averaged for all lanes) is defined here as 2.0 seconds. The value represents a *mean* minimum with original measurements coming as low as 1.0 seconds. Also the region of inflexion where the minimum headway is achieved appears to be **in the range** of 30–40mph.

Figure 30 shows the equivalent data for the offside lane only. Here the minimum mean time-headway achieved at the inflexion points (40–50mph) is around 1.6 seconds. Again this value represents a mean minimum, with original measurements coming in as low as 0.8 seconds. Of course there are questions surrounding the level of accuracy attainable from the loop detectors, but these values are consistent with other recent observations, for example, Brackstone, Sultan and McDonald (2002) where 48% of measured headways were below 1.0 seconds.

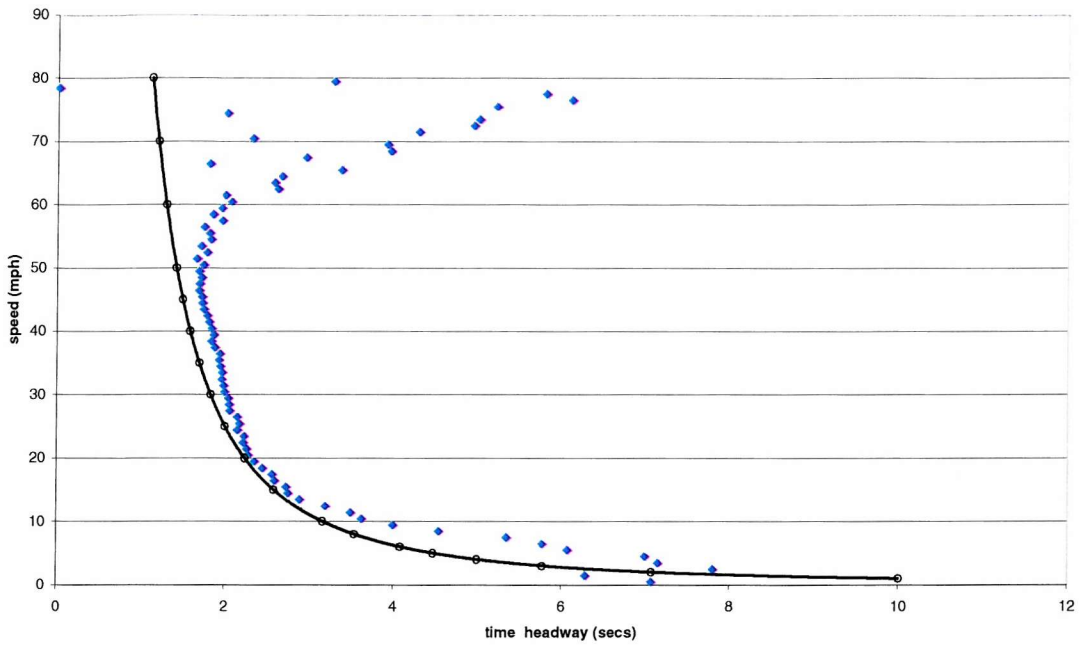


Figure 30 – Speed against time-headway for 9 days at Loop 4902A (offside-1 lane only)

Brackstone, Sultan and McDonald (2002) also proposed a $1/\sqrt{v}$ type relationship for time-headway measurements. A sample curve of this type has been overlaid in Figure 30. The overall profile matches well at speeds of up to around 50mph, after which the greater level of scatter in headways moves the average points further and further to the right.

Looking at Figure 31 and Figure 32 it can be seen that each of the lanes approaches its maximum speed at a different rate. This reflects the distribution of vehicles by type across the lanes, but also contains important information about the 'desired speed' of drivers within each lane. Accurate knowledge of these values would greatly assist the development of microscopic models.

Another interesting point of observation is the fact that the divergence appears above the 40mph mark. This may be indicative of the variable speed control present on this particular motorway, suggesting that vehicles do conform to a control environment, but are seeking to achieve varying levels of acceptable speed once control is lifted.

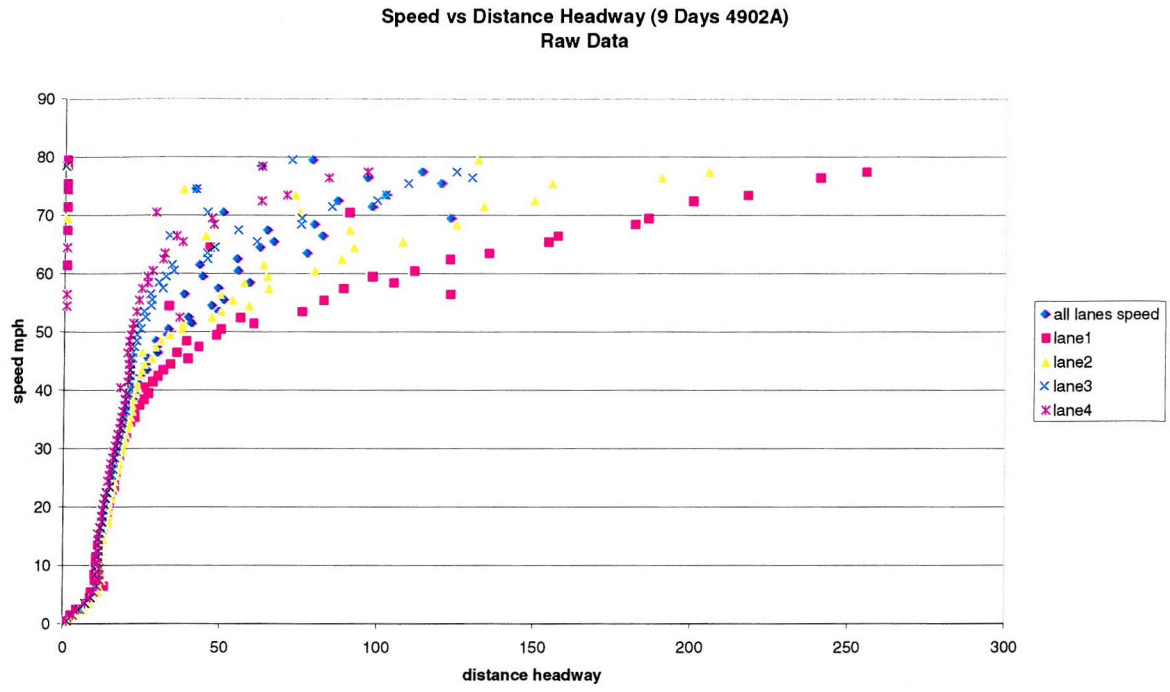


Figure 31 – Speed against distance-headway (raw data for all lanes)

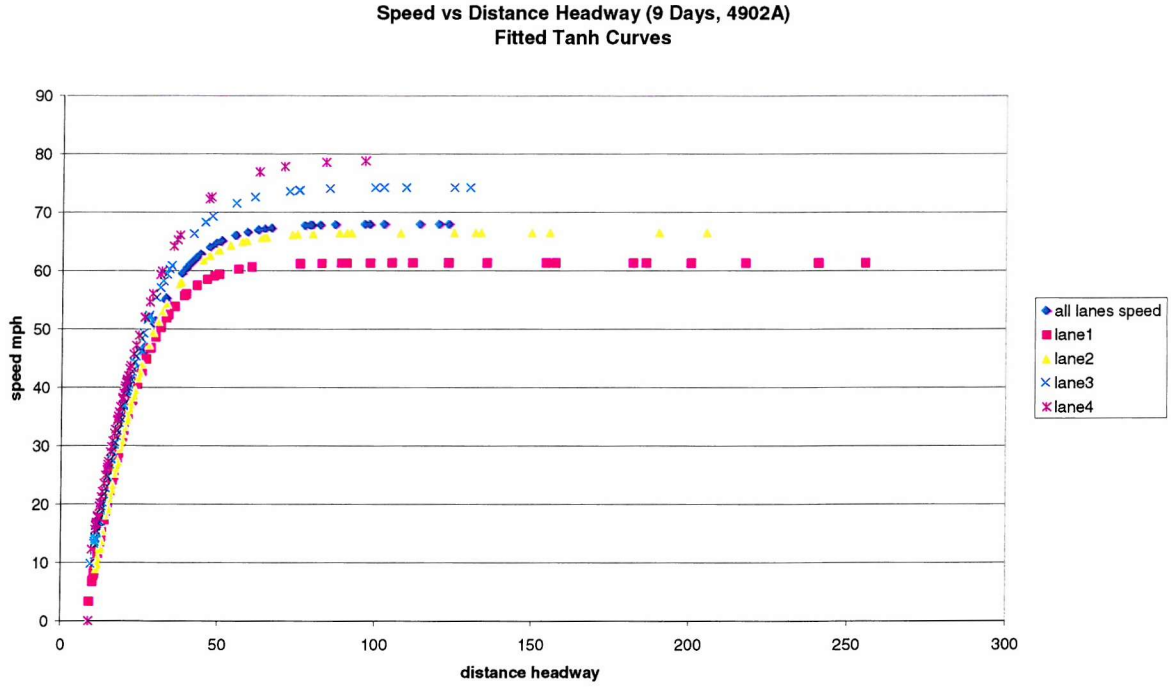


Figure 32 – Speed against distance headway (fitted curves for all lanes)

5.4.3 Old models in a modern context

Taking the speed-headway investigation a step further, various forms of curve (as suggested by Holland) were tested against the mean observed values for speed and distance headway. The 'goodness of fit' of these curves was tested using a least-squares approach. The sum:

$$\sum_{i=1}^{80} (\text{actual}_{v_i} - \text{calculated}_{v_i})^2$$

was formed, where calculated v_i is the speed applicable at the actual headway according to the suggested relationship. This raised a number of questions regarding 'effective speed-headway functions'. It was found that for every case except the inside lane, a relationship of the form:

$$\text{speed} = v_{max} \tanh\left(\frac{\lambda}{v_{max}} \times (\text{headway} - \text{headway}_{min})\right)$$

as suggested by Bando *et al.* (1995) gave a better fit, with the closest match occurring for the fast lane. Figure 33 illustrates this point.

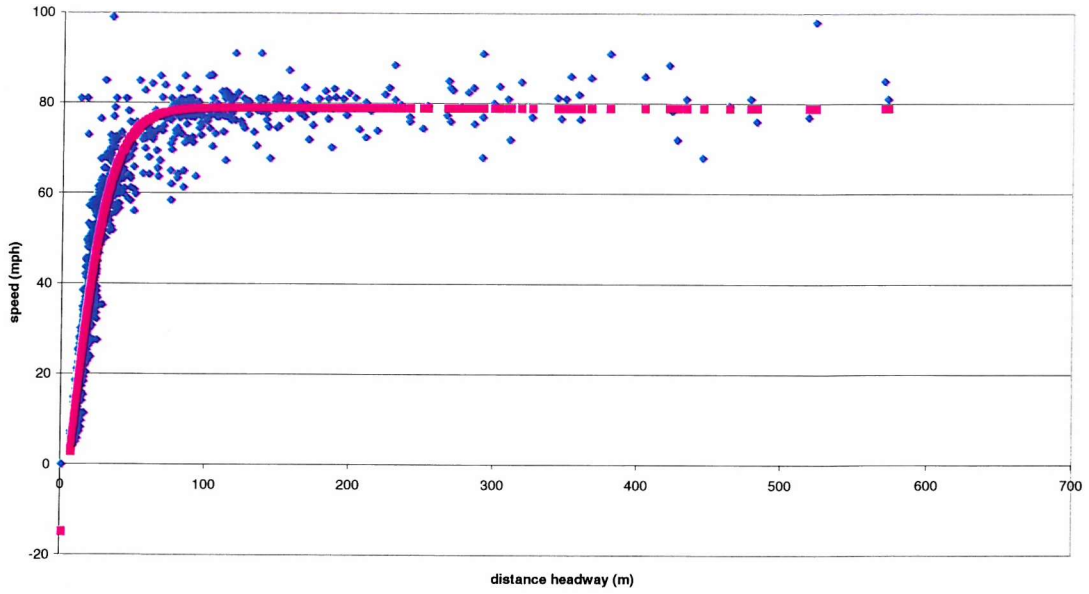


Figure 33 – Speed-headway data for offside lane, with fitted hyperbolic tan function

However, the *slow lane* conforms more closely to an exponential relationship, namely:

$$\text{speed} = v_{max}\left(1 - \exp\left(\frac{\lambda}{v_{max}} \times (\text{headway} - \text{headway}_{min})\right)\right)$$

as suggested by Newell (1961). Looking at the date of this work it may be reasonable to assume that traffic behaviour and vehicle performance on motorways in that era was comparable to traffic groups within the slow lane today (over 40 years later).

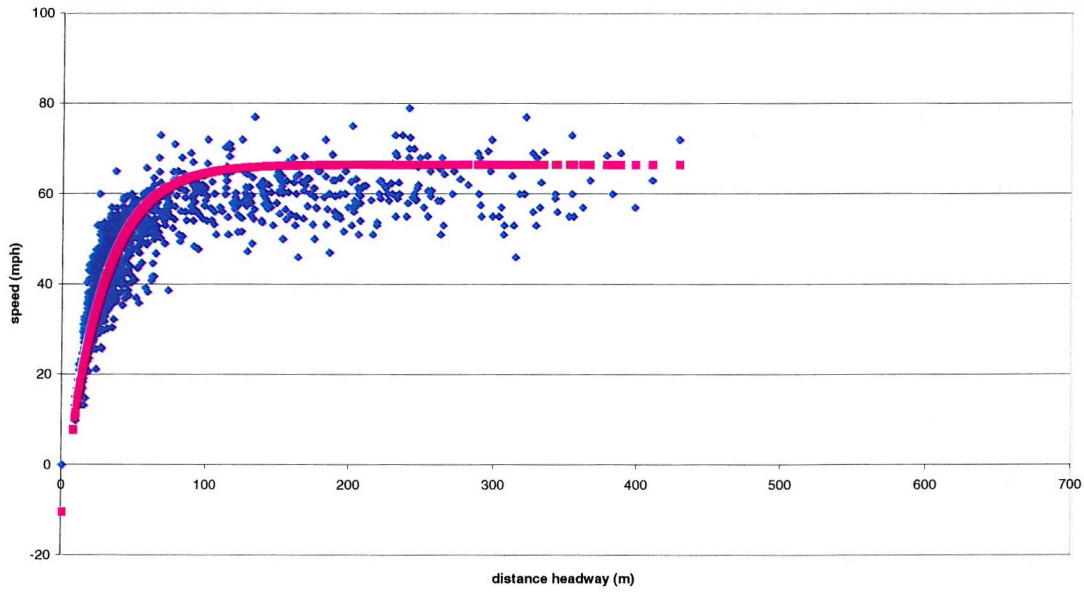


Figure 34 – Speed-headway data for nearside lane, with fitted exponential function

Observed behaviour on the nearside lane does indeed suggest a certain degree of isolation from other lanes, possibly due to its status as the 'slow lane' and the larger concentration of heavy goods vehicles. For the offside lane the sharper gradients within the curve point to a consistent acceptance of lower headway at equivalent speeds.

The range of behaviour across lanes is further emphasised by the fact that intermediate lanes are modelled better on some days using one curve, and on other days using another curve, again depending on the vehicle composition

within the lane. This can be taken further by questioning the validity of constraining the function to the origin. Indeed the IVD points to a minimum headway at zero speed (indicating that a gap is still evident even between so-called 'bumper to bumper traffic').

Further investigation of this is warranted, especially in testing the performance of the classical car-following models such as listed by Holland (1998) using a new speed-headway function calibrated to the M25 IVD. However, this is beyond the remit of the current thesis.

5.4.4 Mapping headway to occupancy

Having considered the nature of the speed-headway function it is now necessary to consider its application to the present study. The primary data source being investigated through the probability model is the MIDAS occupancy value, (the proportion of a one-minute interval for which a loop is active, given as a percentage). A simple transform is possible although the output generates a significant amount of variance about the mean.

It is possible to derive a simple mapping for distance headway values calculated from the IVD and the detector occupancy readings. In one minute an average number of vehicles (N) will move a total distance (D). This value comprises the sum of all these vehicle lengths and the sum of the distance between them. Occupancy can therefore be rationalised to ~ (100L)/(L+H)% for an average length (L) and an average vehicle separation (H).

The initial speed thresholds are derived from examination of the plots for speed against time-headway (see Figure 29 and Figure 30). The process for deriving the occupancy thresholds is a little more involved. A closer examination of the speed-headway curves shows that speeds start falling when the corresponding distance headways reach values of around 25m (see Figure 31). The average vehicle length is different in each lane (consider the vehicle composition). Assuming an average length of 5m gives ~ (100×5)/(5+25)% which is 17%. From Figure 35 it can be seen that this maps to around 20% occupancy. A similar calculation for an average of 3.5m translates to around 16% occupancy. These values could therefore be

considered as the *lower bounds* for the thresholds. In other words a data point with speed of 40mph and occupancy of around 20% is an approximation of the transition from free flow to synchronised flow. The upper bound is harder to determine, although inspection of the speed-headway plots would again suggest that speeds below 20mph are representative of flow-breakdown, and the corresponding distance headway of 10–15m translates to occupancy of around 30–35%.

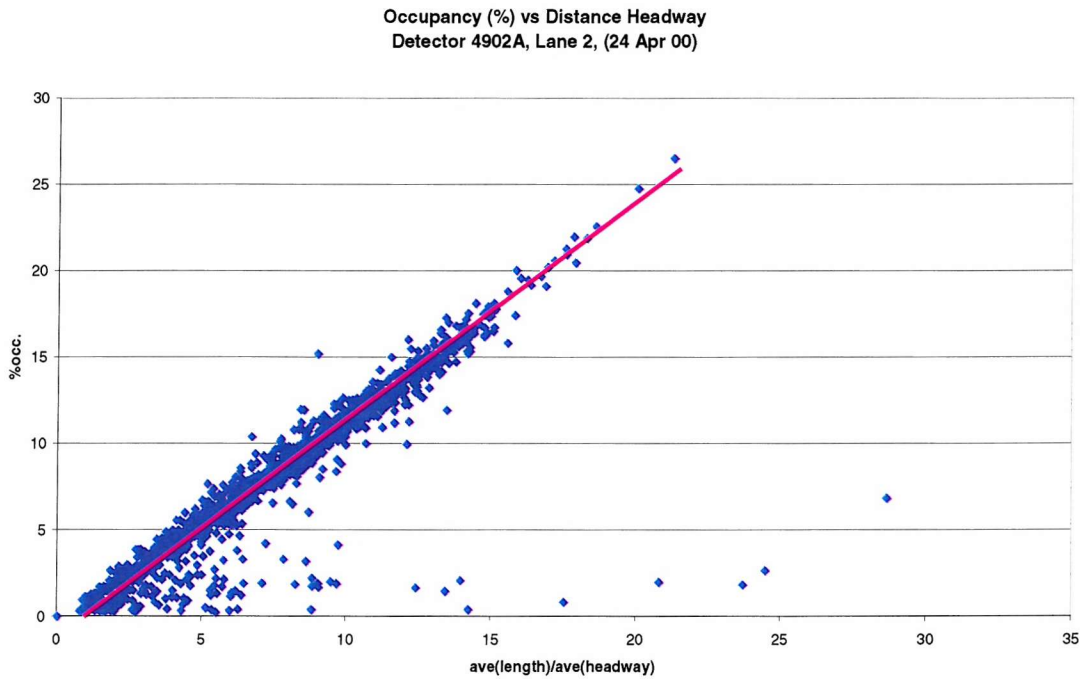


Figure 35 – Mapping calculated occupancy to measured occupancy

5.4.5 Summary for speed-headway analysis

Individual vehicle data was collected from the M25 at sites located near J13 (A30) and J16 (M40), and the relationship between speed and headway was investigated. This particular aspect of the research has yielded some interesting results, particularly on the nature of the curves applicable to different lanes, and the possible influence of vehicle composition.

More importantly, the investigation has identified critical regions on the curves that may provide guidance on identification of the seed-point (or at least traffic conditions which are characteristic of a seed-point) during the detection

phase. These regions have been translated using a basic method for mapping headway to occupancy. Analysis has led to a more informed first pass at deriving thresholds for the detection algorithm to operate. However, it remains prudent to vary the parameters in order to test for sensitivity to the definition.

5.5 CONCLUSIONS FOR SEED POINT DETECTION

This chapter has shown that it is possible to identify the location of seed-points by classification using traffic characteristics. The development of the main algorithm for seed-detection started through investigating the performance of incident-detection algorithms. The California Algorithm differentiated well between incidents and shockwaves, and the McMaster Algorithm provided good consistent detection of shockwaves with little noise in the output.

Investigating these algorithms helped to emphasise the *distinction* between the shockwave body and the shockwave origin point. This led to the conclusion that incident detection alone is not sufficient for seed-point detection, and a new algorithm was developed as part of a semi-automated approach to classify the threshold between low speeds and high traffic density. The current state of software developed by the author specifically for this study (SeeDetect©) has also been presented.

The research then focussed on refining the process through understanding the nature of the speed-headway function for motorway traffic. This was done in both theoretical terms and through the development of an empirical relationship for the M25 using individual vehicle data. The speed-headway analysis provided confirmation that the seed-points lay on the threshold between falling speed and rising occupancy.

The methodology presented here has been used to assemble a database of seed-points along with other critical information contributing to the probabilistic model described in the previous chapter.

6 METHOD APPLICATION AND RESULTS

6.1 STATING THE METHOD

In Chapter 4 the probabilistic model was presented in terms of a Bayesian Equation requiring a prior function, a likelihood function, and a normalising constant. In Chapter 5 there was a discussion on the means by which the specific components for the probability model could be derived, concluding that the seed detection algorithm at the heart of the software would be useful for obtaining the necessary data sets in a consistent manner. Derivation of traffic data from MIDAS Outstations on the M25 Controlled Motorway has already been described in Chapter 3. The SeeDetect© software described in Chapter 5 is capable of reading and processing data in this format.

To help focus on the method being presented in this chapter, the basic steps are summarised below:

- (Step 1) Detecting seed-points within a defined search space and sorting the data;
- (Step 2) Calculating key components for probabilistic equation, and deriving the subsequent posterior matrix;
- (Step 3) Creating a probability map using real detector data and evaluating the predictions against occurrence of shockwaves.

This is the basic methodology under which the results will be presented. In terms of thinking deeper about the method presented, five questions are tackled throughout the chapter as follows:

- If thresholds defining the search space are changed, what effect does this have on the results?
- Is there any benefit in calculating statistical distributions (such as the bivariate normal) if the posterior can be derived from raw data?
- What effect does using an improved prior have on the posterior matrix for new data?
- Can coverage of flow-breakdown be improved by combining output from the probability matrix and the incident detection algorithms?
- Is there a relationship between upstream traffic conditions and the occurrence of a shockwave?

6.2 APPLICATION OF THE METHOD

6.2.1 Detecting the seed-points (Step 1)

The facility for manually selecting seed-points was used in the tuning process to find suitable thresholds for Rules 3–5. Speed and occupancy thresholds were estimated from the analysis presented in Section 5.4.2. But for implementing the method, a decision was taken at the start not to tamper with the automatic selection process. This would ensure that any variation on the part of the detection algorithm from the real world data would be systematic.

Before proceeding with detection it is necessary to make one *important assumption*, that values of speed and occupancy outside a defined range correspond to traffic that is free-flowing or already in flow-breakdown. In other words, the probability of those locations being a seed-point is zero. This assumption is supported by the argument already presented in Chapter 5, but is also numerically acceptable in terms of the posterior matrix (since null values in f_1 will make the actual values of f_0 irrelevant).

The assumption is translated into this methodology by taking the rule base from Chapter 5, (the Seed Detection Algorithm), and defining the basic data set as being all points within the search space satisfying Rules 1–2. For ease of reference, the rule base is listed again here as follows:

1. verification that the occupancy value lies within a predefined range;
2. verification that the speed value also lies within a predefined range;
3. time series comparison of speed and occupancy, to verify a rise in occupancy and a simultaneous fall in speeds;
4. additional check on the numerical proximity of speed and occupancy using these particular units, to assist with classification;
5. final check for direction of the gradient on speed and occupancy profiles, (to distinguish a seed-point from the edge of a shockwave).

Figure 36 shows how the search space is defined using Rules 1–2 from this algorithm. The 'predefined range' was chosen as (24<v<38) and (24<k<35), this is referred to as Automatic Parameter Set 1 (or APS1). By definition all seed-points in this period will be a subset of the cluster shown in the figure.

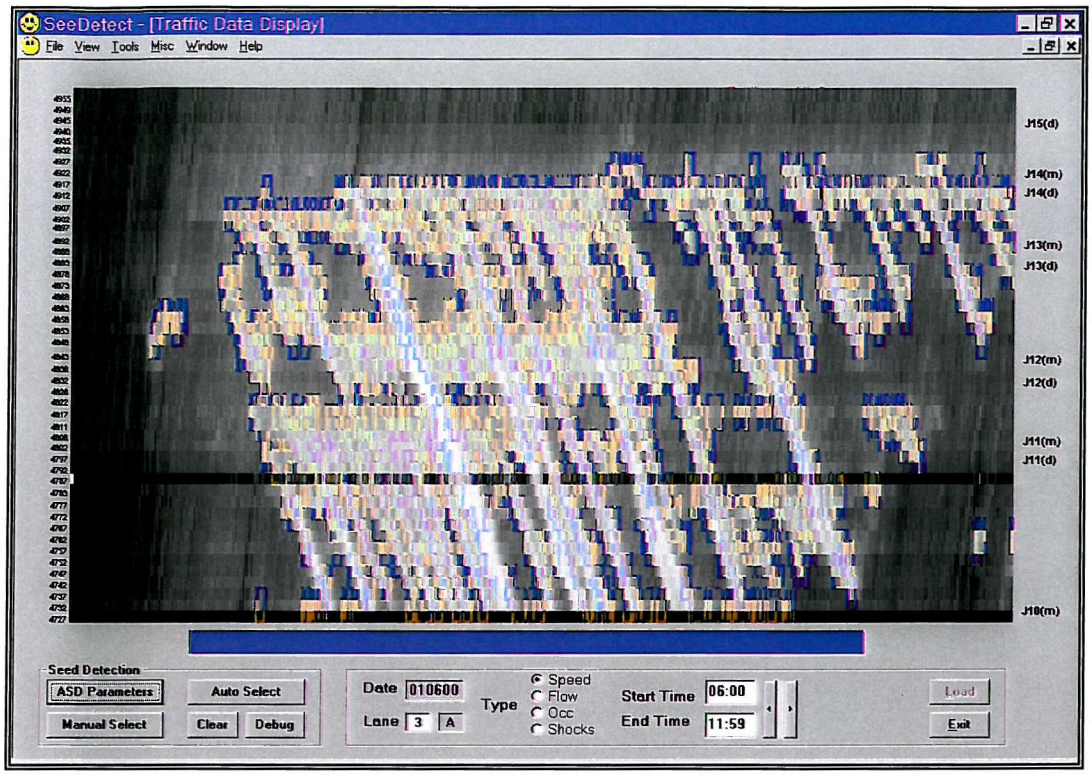


Figure 36 – Defining the seed-point search space by applying Rules 1–2

Thus the selection of automatic search parameters (the thresholds for speed and occupancy) is influential on the results. Later in the chapter the sensitivity of results to variations in this parameter set is briefly examined. The SeeDetect© software was used to search over ten days (from 00:00 to 23:59 on each day and covering loops 4727A to 4955A). As described in Section 5.3, the data records created include information from the eight points surrounding the qualifying centre point.

Figure 37 presents a screenshot from the SeeDetect© software for APS1. The automatic detection process has identified a range of points that satisfy all five conditions in the seed detection algorithm (see Section 6.1). For the most part, these points are clustered around the start of the shockwave (in regions of less pronounced variation in the traffic parameters), as opposed to the shockwave body itself – which is in itself encouraging. On some days the shockwaves form clearly punctuated patterns with drivers picking up speed between successive shockwaves, whereas on other days speeds in this intervening period can be similar to those experienced in the proximity of a seed-point.

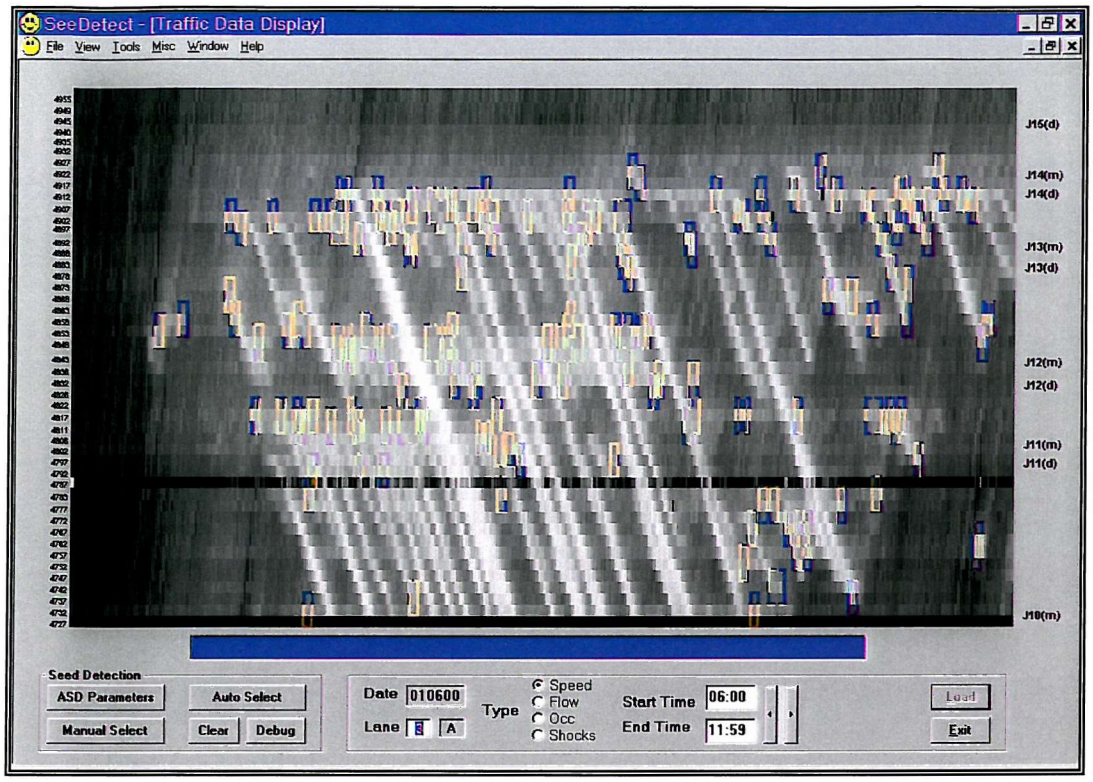


Figure 37 – Selecting seed-points within the search space using Rules 1–5 and Parameter Set #1 (01 June 2000)

Once the two data sets (the search space and the qualifying seed-points) had been obtained using the SeeDetect© software, the data was then sorted. The search space contained over 10,000 records but this was still significantly less than calculating for every single cell (which would have been of the order $1440 \times 47 \times 30 = 2{,}030{,}400$ records)! A subset of records satisfying Rules 1–5 (the seed-points) was then extracted from this data set . The extraction was done using various commands within MS Excel. By combining the entries for pntDate, pntLoop, and pntTime a string unique to that record was obtained and appended. Once the smaller 'seed-point' set was obtained a similar unique marker was used, and the two data sets were compared using the MATCH command. Duplicate records could then be deleted after sorting on the appropriate column.

6.2.2 Calculating the probability matrix (Step 2)

The prior function was initially calculated as the proportion of qualifying points to the total data set. In this case there were 1,364 'seed-points' out of a

possible 11,638 pairs within the search space. On the basis of Section 4.3.3 the prior $\pi(\theta=1)$ is simply a uniform value of 0.12. In real terms this prior asserts an initial belief that each pair of values within the search space (v, k) has an equal probability of being a seed-point. The difference (11,638 – 1,364 = 10,274) is therefore the total number of pairs not qualifying as seed-points.

The likelihoods $f_1 = f(q \mid \theta=1)$ and $f_0 = f(q \mid \theta=0)$ were obtained by examining the distribution of speed and occupancy within the 1,364 pairs and the 10,274 pairs respectively. Matrices were set up and the frequency function within the spreadsheet was applied to a combined string to derive the number of times each unique pair occurred. The pair (26, 30) was denoted as '26030' and so on.

Figure 38 presents the results for f_1 as a surface plot based on discrete data, with a peak at (31, 26). Figure 39 presents similar information but for f_0.

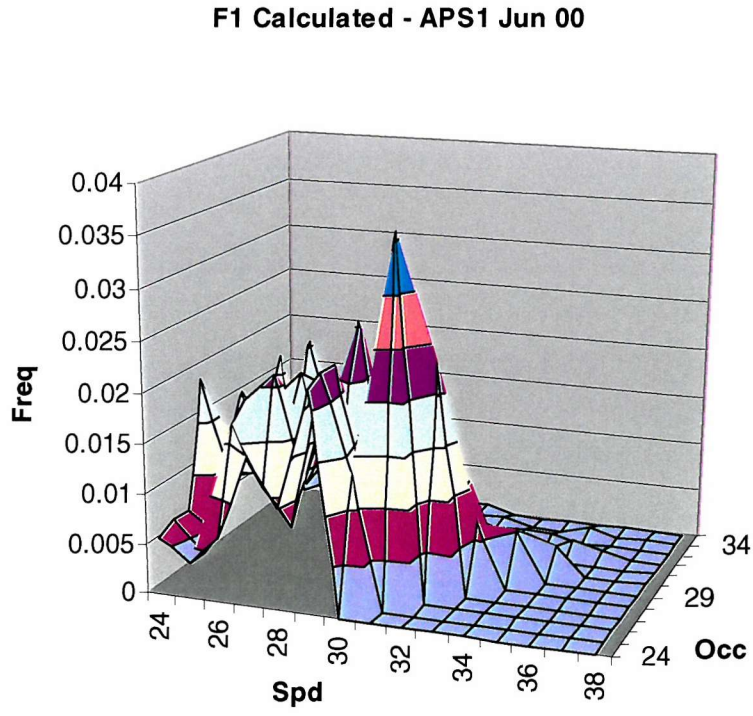


Figure 38 – Frequency distribution for f_1 showing abnormal peak for (31, 26)

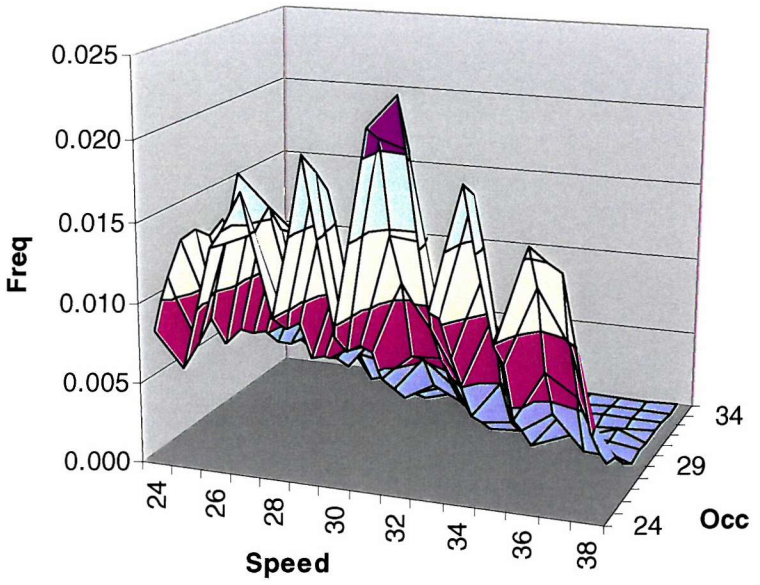


Figure 39 – Frequency distribution for f_0

Recall the equation for posterior distribution from Section 4.3.5

$$f(\theta=1|q)=\frac{f(q|\theta=1)\cdot\pi(\theta=1)}{[f(q|\theta=1)\cdot\pi(\theta=1)]+[f(q|\theta=0)\cdot(1-\pi(\theta=1))]}$$

or in simpler notation

$$f(\theta=1|q)=\frac{f_1\cdot p}{[f_1\cdot p]+[f_0\cdot(1-p)]}$$

All the component parts have been derived, so a new posterior matrix can be calculated for (24< v <38) and (24< k <35). Figure 40 gives the surface plot for the outcome. The surface height at each coordinate translates into the probability of the values at that coordinate being associated with a seed-point. Unlike the previous two figures, this is **not** a cumulative frequency chart. Table 2 gives the corresponding values and highlights the peak value of 0.42 at (30, 30).

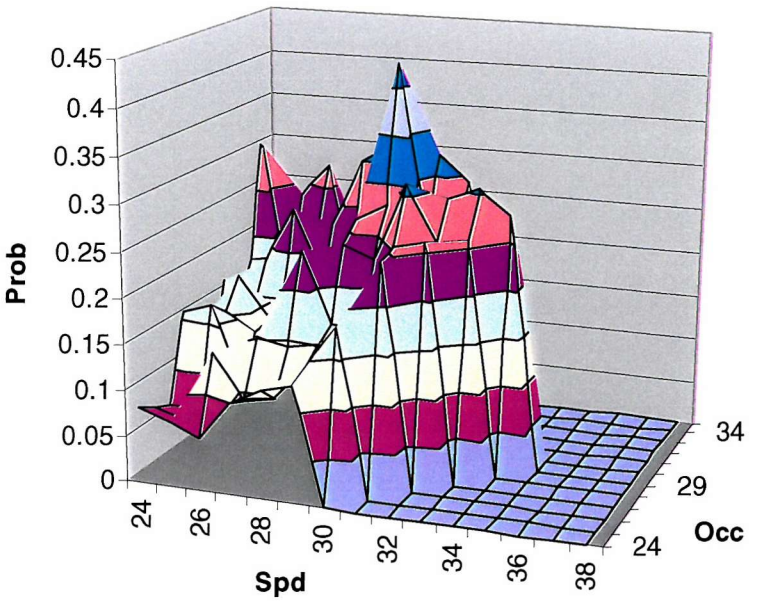


Figure 40 – Posterior distribution derived from frequencies for f_1 and f_0

Occupancy \ Speed	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
24	0.08	0.07	0.06	0.10	0.12	0.13	-	-	-	-	-	-	-	-	-
25	0.07	0.06	0.14	0.09	0.10	0.14	0.18	-	-	-	-	-	-	-	-
26	0.06	0.07	0.12	0.15	0.14	0.15	0.17	0.25	-	-	-	-	-	-	-
27	0.16	0.17	0.08	0.13	0.24	0.16	0.16	0.25	0.26	-	-	-	-	-	-
28	0.14	0.15	0.13	0.17	0.18	0.24	0.23	0.31	0.25	0.26	-	-	-	-	-
29	0.10	0.18	0.14	0.19	0.16	0.26	0.28	0.30	0.29	0.30	0.28	-	-	-	-
30	0.17	0.13	0.14	0.13	0.20	0.20	**0.42**	0.25	0.25	-	-	-	-	-	-
31	0.12	0.18	0.25	0.17	0.17	0.19	0.32	0.32	0.29	-	-	-	-	-	-
32	0.17	0.18	0.20	0.22	0.30	0.32	-	0.21	-	-	-	-	-	-	-
33	0.30	0.12	0.21	0.19	0.26	0.18	-	-	-	-	-	-	-	-	-
34	0.14	0.24	0.27	0.13	-	-	-	-	-	-	-	-	-	-	-
35	-	0.12	0.21	0.15	-	-	-	-	-	-	-	-	-	-	-

Table 2 – Values for posterior distribution shown in Figure 40 (with peak value in bold)

6.2.3 Creating a probability map (Step 3)

Information provided in the probability tables is simple and pertinent, that for a particular coordinate pair (v, k), the probability of a shockwave occurring is $P_{(v, k)}$ based on patterns implicit within the particular data set examined.

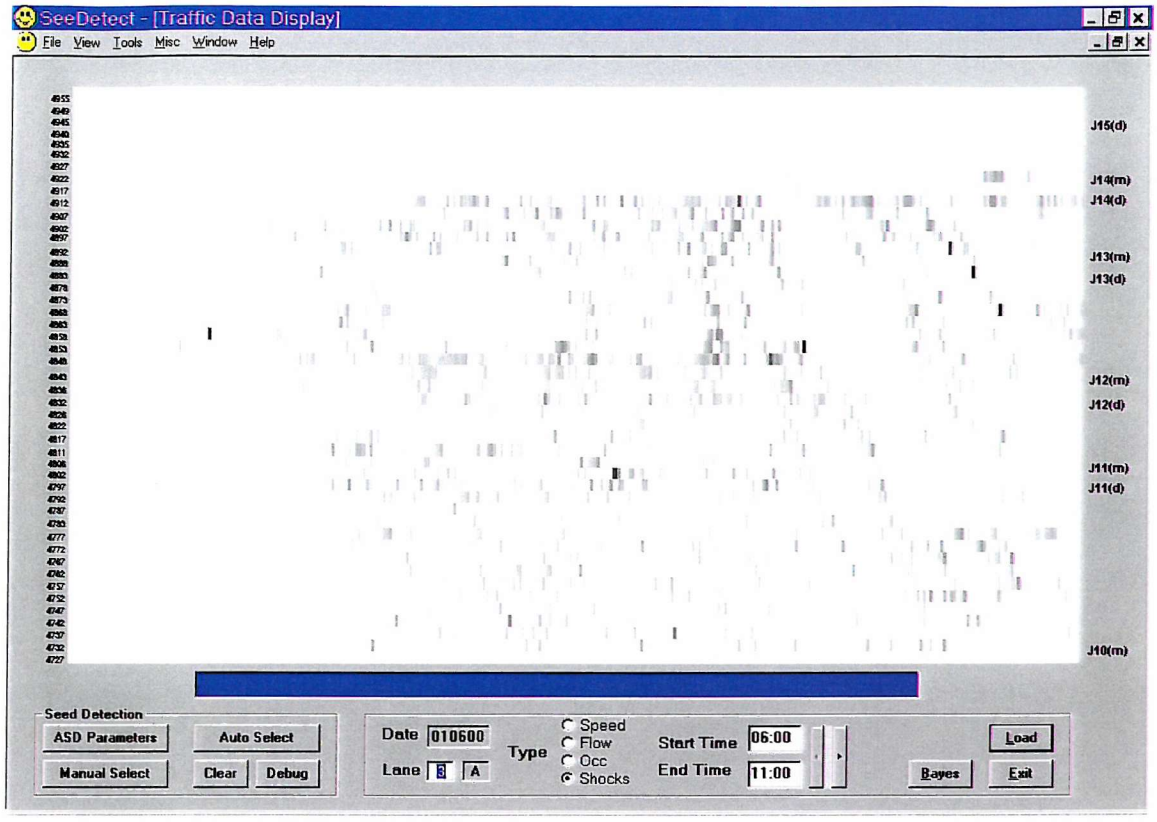
The posterior distribution was imported by SeeDetect© and used to produce 'probability maps' by scanning traffic data in the position-time array and assigning the appropriate value for each combination of speed and occupancy found. Figure 41 illustrates this for 01 June 2000 (as compared to the actual traffic data for that period). The values displayed are from the posterior distribution for 01–10 June 2000. Plots for additional days are found in Appendix E.

These plots offer some confidence that the predictions based on the posterior distributions are in the right range. But in order to evaluate the accuracy (and therefore the usefulness) of the probability maps, it is necessary to focus on the values generated at the individual detector.

Figure 42 shows the minute by minute values of the probability index, compared with the values for speed and occupancy. Some caution is required with the interpretation of this plot, since the variation at a particular detector does not indicate the severity of a shockwave (for this information we need to look at upstream detectors).

A simple implementation of the seed-detection algorithm could be used to watch the data stream at each detector, without the need for any probability analysis. As seen in Figure 28, one key characteristic of the seed-point was the fact that speed would be falling whilst occupancy would be rising, and that numerically they approached a common value (or crossed within a certain threshold).

Figure 43 shows the fluctuations in speed and occupancy between 08:00 and 09:00 for the same detector on the same day. Here the distinctiveness of the seed-point pattern is not so apparent, and although the probability index is still tracking the changes, it is of no practical value.



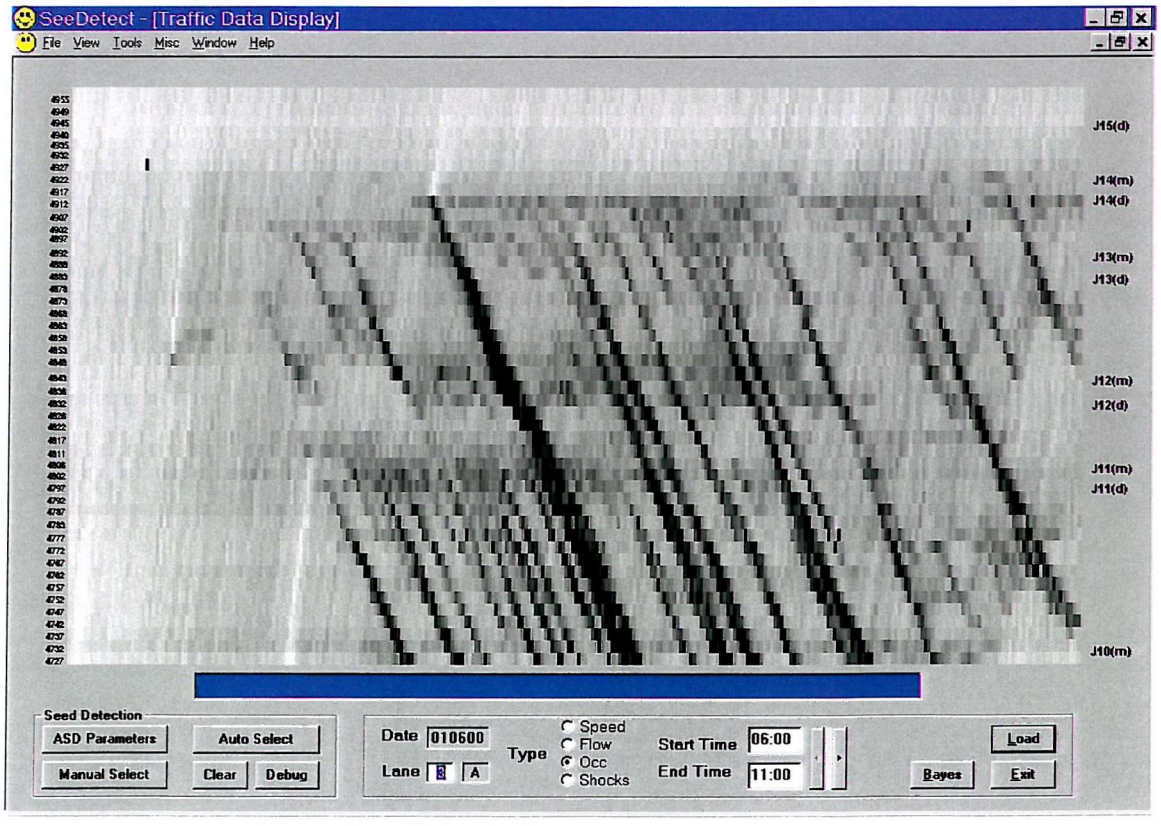


Figure 41 – Probability map and corresponding traffic data for 01 Jun 00 using June 00 Posterior

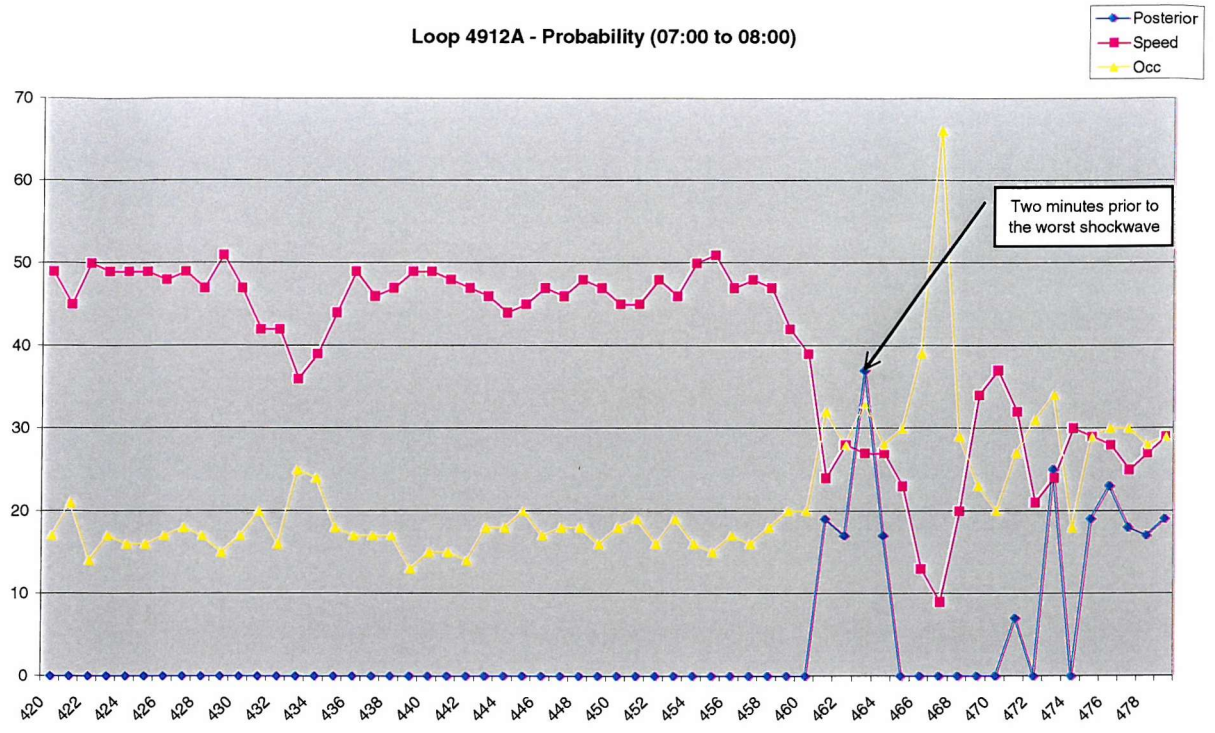


Figure 42 – 'Predictions' at detector 4912A (07:00 to 08:00 on 01 June 2000)

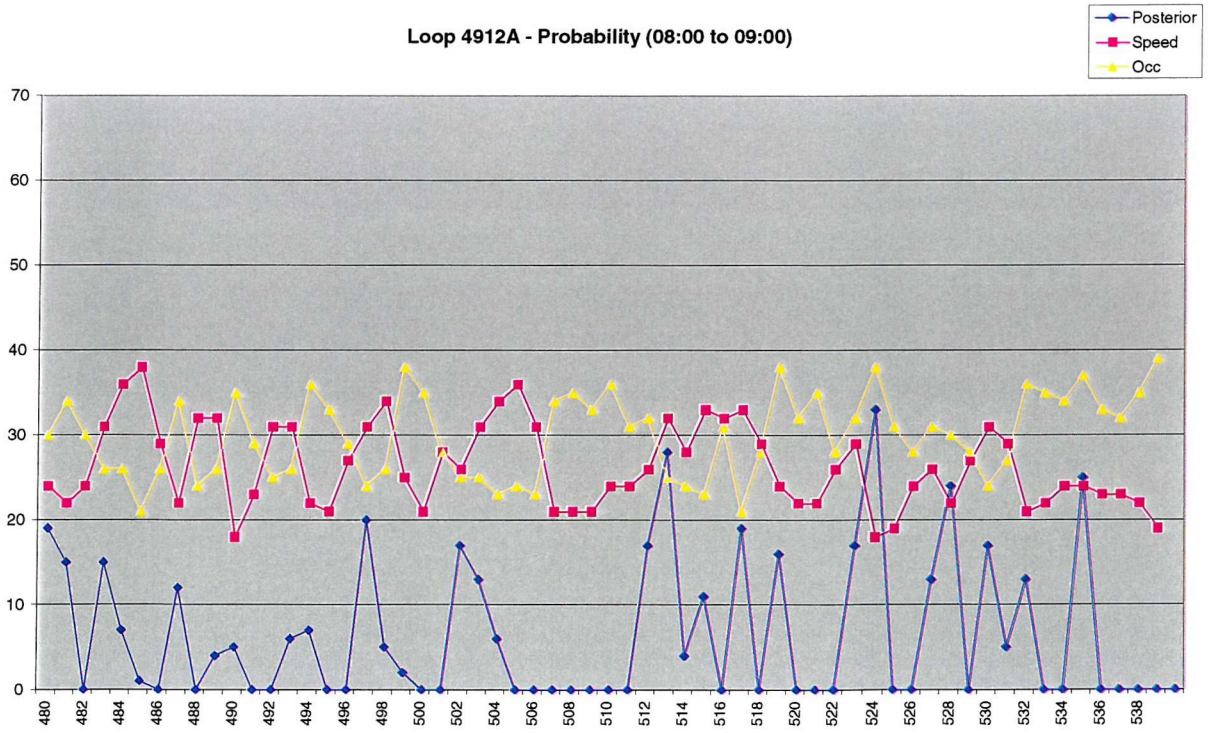


Figure 43 – 'Predictions' at detector 4912A (08:00 to 09:00 on 01 June 2000)

Watching the values of speed and occupancy alone would not provide sufficient information about the possibility of a shockwave occurring, since we know from the existence of a large data set for f_0 that a particular combination of these values would not necessarily lead to a shockwave. It might be possible to refine the algorithm at the heart of the software further (along with the corresponding parameters) in order to maximise the accuracy of seed-point detection. This might be considered sufficient for certain traffic operations. But in general, guidance provided by the probability map will be necessary to make sense of the data detected, and effort should be concentrated on refining the map itself.

The analysis in this thesis has concentrated on a particular definition of seed-points, thereby encompassing regions of minor variation in traffic parameters which hold similar characteristics to seed-points but do not lead directly to shockwaves. But a derivative of this work might tune the algorithm to concentrate on shockwaves only, thereby producing a different posterior distribution.

This analysis also underscores discussion at the start of the thesis about the subtleties of detecting a seed-point. The probability index never goes much beyond around 50%, reflecting the fact that although the seed-point has been characterised to such an extent that we can identify it, there are still many occasions when the same speed and occupancy conditions will not lead to a shockwave. The main advantages of following through on producing the posterior distribution (rather than depending solely on the detection algorithm) are as follows:

- If dealing with detectors that are only giving intermittent readings, it is not necessary to depend on consecutive readings (as used in the algorithm itself), only on the speed and occupancy readings obtained at that particular minute of interest.
- It is possible to obtain an indication of the likelihood that there will be a shockwave, rather than just assume that a 'successful identification' will turn out to be a shockwave.
- The performance of the algorithm itself can be tested through a feedback process, so that the impact on both f_1 and f_0 of any changes made to the detection logic can be explored.

6.3 FURTHER INVESTIGATIONS

6.3.1 Preliminary test of threshold sensitivity

It is important to understand the sensitivity of results to variation in thresholds which define the search space. Optimisation is achieved when thresholds are varied so that an increment brings no additional values to the data set for Rules 1–5 (the data set for Rules 1–2 is much broader). This is a time-consuming activity and would need to be conducted over a number of days in order to establish an 'acceptable range'.

For the purpose of this study only one variation has been presented, on the basis that the selection of thresholds has been approached from the top-down (through speed-headway analysis) not the bottom-up (numerical analysis of sample size). Changes were made only to the thresholds for the first two rules, such that the search space under APS2 was (15<v<38) and (15<k<35).

During the same period analysed in Section 6.2 there were 1734 pairs qualifying as seed-points (an additional 370). The increase in search space yielded a total of 31,219 pairs, of which 29,485 were not seed-points. In proportional terms a 168% increase in search space yielded a 27% increase in identified seed-points.

The comparison can also be seen in Figure 44 (showing APS1) and Figure 45 (showing APS2). The left column shows the frequency distributions in each case with the boundary for f_1 shown clearly in Figure 45. Choice of thresholds will therefore have two effects on the results:

- The initial estimate of the prior (as a ratio) will be affected by the quantity of points allocated to each data set. This becomes less relevant once the process for estimating the prior is improved (see Section 6.3.3).
- The search space must include all points which qualify as f_1 although it is less important (in terms of calculating the posterior) if the search space is smaller than f_0. This is because when $f_1 = 0$ for a particular set of (v, k) then $f(\theta=1|q)$ is also zero, (see Figure 45 in particular).

6.3.2 Bivariate Normal Distribution

Given that the detection algorithm functioned on two variables (v, k) and that there was a possible relationship between them, it was noted in Section 4.3.4 that a bivariate normal distribution would be most appropriate for describing the data sets. Statistical distributions are usually sought out in order to assist with complicated calculations. Because the detected values for speed and occupancy are discrete, it has been possible to obtain the corresponding likelihoods using frequency alone. However, there are a number of benefits in attempting to create an equivalent distribution.

Comparative plots for f_0, f_1 and the posterior distribution are given in Figure 44 and Figure 45. The right column in each figure presents a bivariate normal distribution based on statistical parameters calculated for the data set. Further examples for June 2001 and September 2001 can be found in Appendix E.

Although the bivariate normal distribution does not capture the peak very well, it provides an adequate description of curve shape. In terms of application, the bivariate normal function offers considerable flexibility. The initial data set is described by a set of statistical parameters, namely the means, standard deviations, and covariance (μ_k, μ_v, σ_k, σ_v, ρ_{kv}). If new data is collected (say for the following month), that data can then be described using similar parameters. Both sets of parameters can then be used to calculate a new combined value. The following relationship should be considered:

$$f(\theta \mid q_a) \propto f(\theta) \times f(q_a \mid \theta)$$

$$\begin{aligned} f(\theta \mid q_a, q_b) &\propto f(\theta \mid q_a) \times f(q_b \mid \theta) \\ &\propto f(\theta) \times f(q_a \mid \theta) \times f(q_b \mid \theta) \\ &\propto f(\theta) \times f(q_a, q_b \mid \theta) \end{aligned}$$

Traffic variable sets q_a and q_b represent data collected in different time periods (for example, in consecutive months). The above relationship can therefore be applied since q_a and q_b can be assumed as independent. If the objective is to accumulate a large data set then the practicality of this approach for incremental development becomes evident.

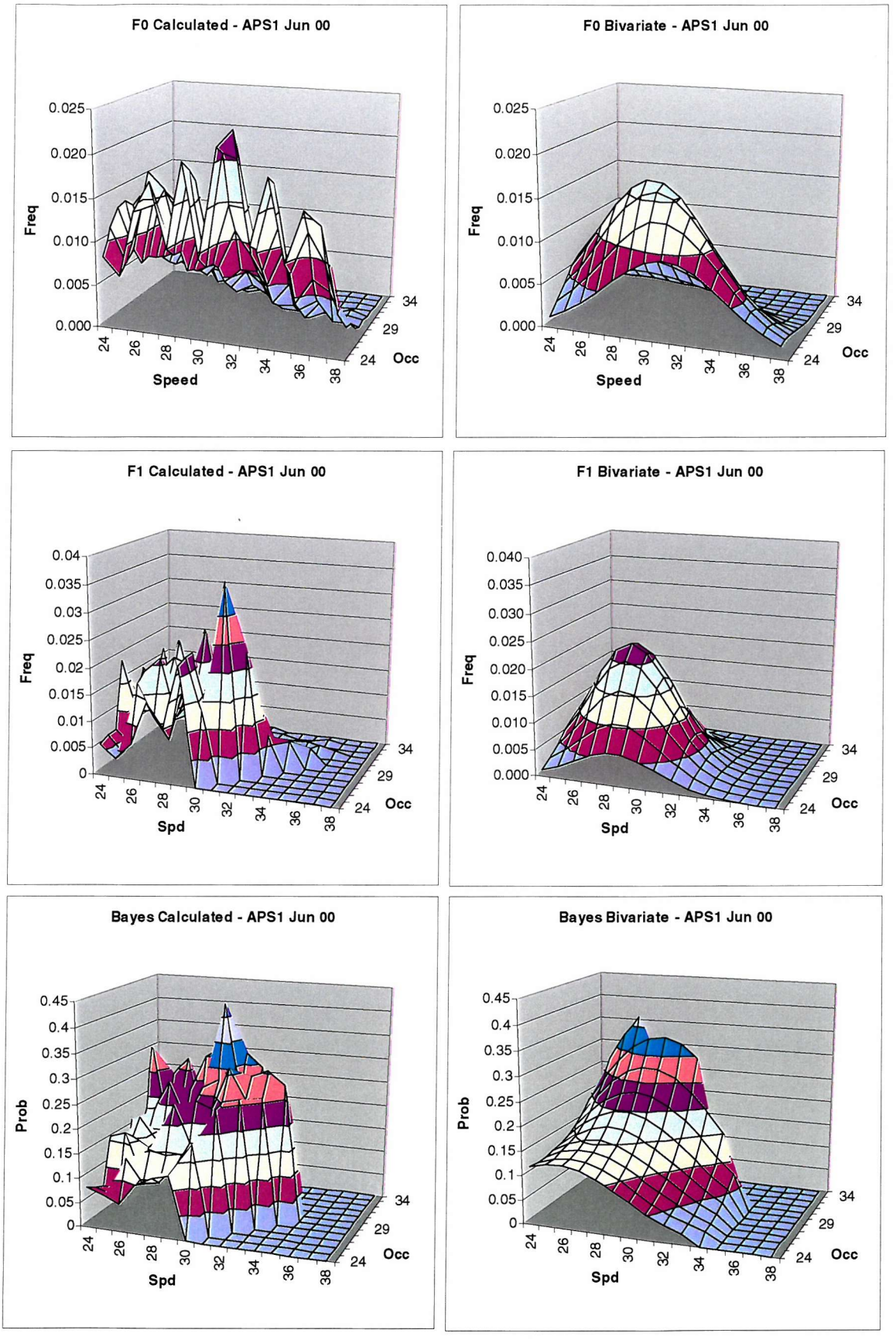


Figure 44 – Data and statistical distributions (f_1, f_0, and posterior) for APS1

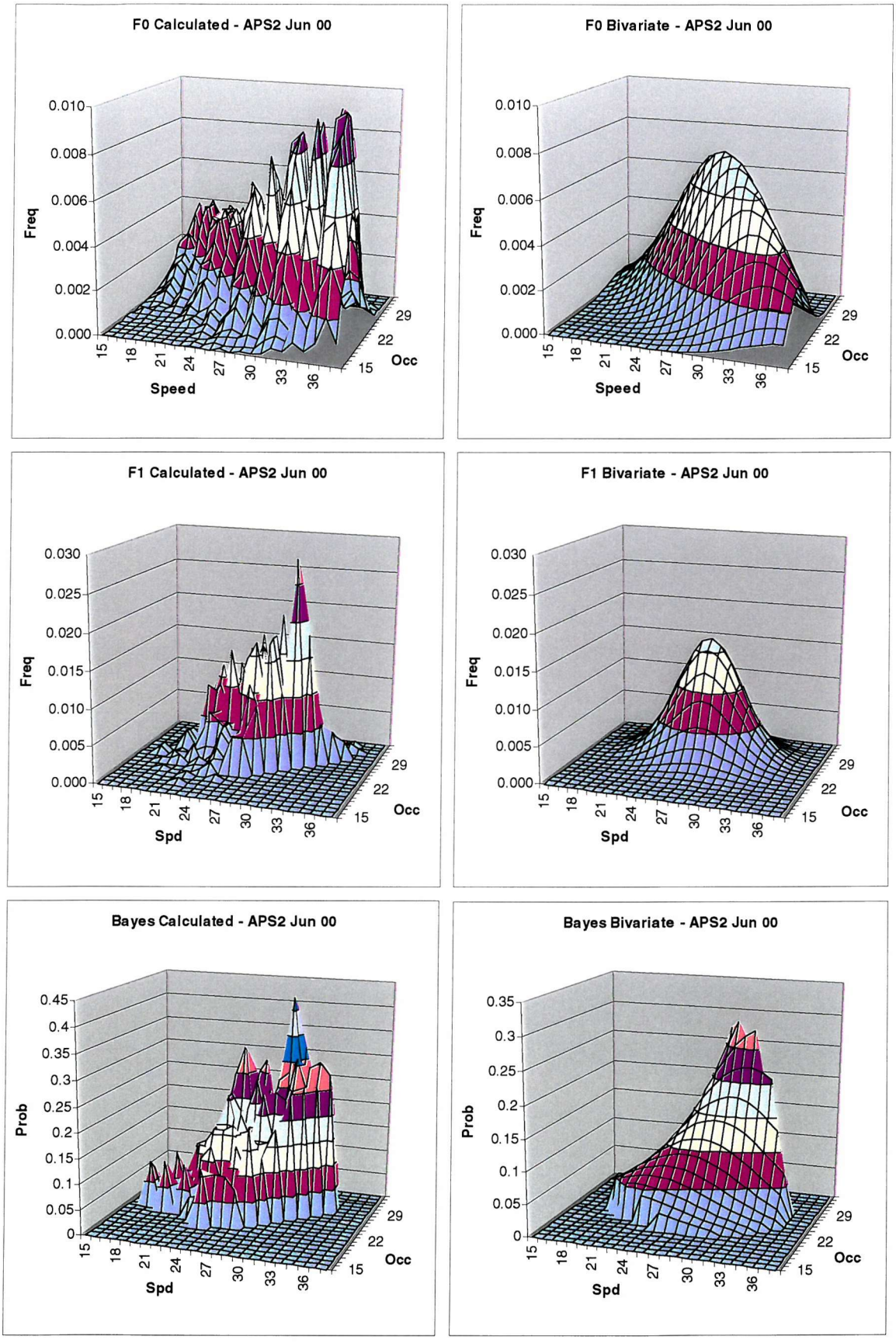


Figure 45 – Comparative data and statistical distributions for APS2, showing a fuller range for f_0 and the values for f_1 clearly bounded

Table 3 lists the statistical parameters associated with Figure 44 and Figure 45. For each parameter set (APS 1 and 2) the mean speed is slightly lower for the seed points (Rules 1–5), and the mean occupancy is slightly higher. For comparison, further values are listed in Table 4 and Table 5, the corresponding graphics are given in Appendix E.

Table 3 – Parameters for bivariate normal distribution, derived from June 2000 data

		μ_v	μ_k	$(\sigma_v)^2$	$(\sigma_k)^2$	cov (v,k)	$\rho_{v,k}$	sample
APS1	Rules 1-5	27.96	28.30	6.91	6.92	-0.89	-0.13	1364
	Rules 1-2	29.42	26.85	15.38	6.99	-4.15	-0.40	10274
APS2	Rules 1-5	26.80	28.15	11.45	7.98	-0.78	-0.08	1734
	Rules 1-2	29.06	24.76	39.14	23.13	-21.83	-0.73	29485

Table 4 – Parameters for bivariate normal distribution, derived from June 2001 data

		μ_v	μ_k	$(\sigma_v)^2$	$(\sigma_k)^2$	cov (v,k)	$\rho_{v,k}$	sample
APS1	Rules 1-5	27.54	28.05	6.49	7.41	-0.36	-0.05	786
	Rules 1-2	29.47	26.68	15.91	6.32	-4.06	-0.40	4751

Table 5 – Parameters for bivariate normal distribution, derived from September 2001 data

		μ_v	μ_k	$(\sigma_v)^2$	$(\sigma_k)^2$	cov (v,k)	$\rho_{v,k}$	sample
APS1	Rules 1-5	27.57	27.72	6.38	6.44	-0.44	-0.07	3698
	Rules 1-2	28.74	26.62	14.38	6.39	-3.53	-0.37	25677

By converting the bivariate normal distributions for f_0 and f_1 into contour maps it is possible to observe the level of correlation present. Figure 46 illustrates the frequency distribution for the f_0 (satisfying Rules 1–2 only) using APS2 for June 2000.

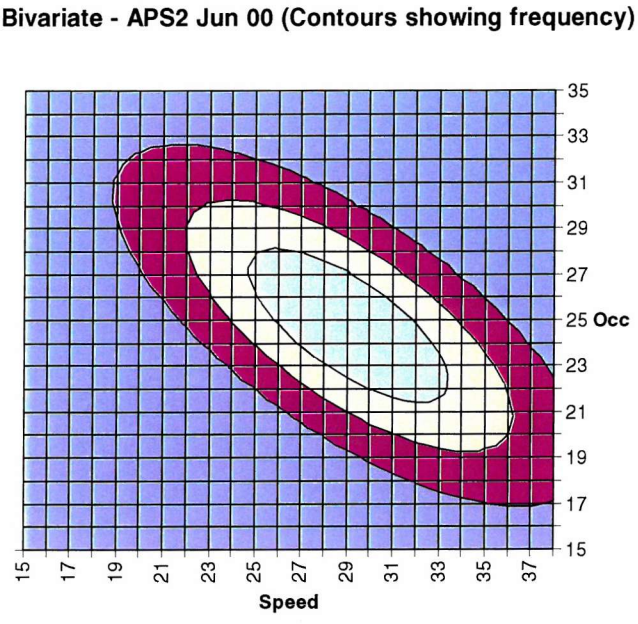


Figure 46 – Contour maps for the distribution of f_0 for APS2 using June 2000 data, the shape of the contour indicates correlation (and therefore an implicit relationship) between the parameters

The change in shape between APS1 and APS2 is simply due to the additional data bringing a change in the variance and correlation of the total set. The position of the peak is determined by the coordinates of the mean values, the spread of the data is measured by the variance. In APS2 the correlation value is -0.73 suggesting a strong relationship between speed and occupancy. The implication is that there is an overall tendency for speeds to fall as occupancy rises.

Figure 47 gives similar information but for the f_1 set (satisfying Rules 1–5).

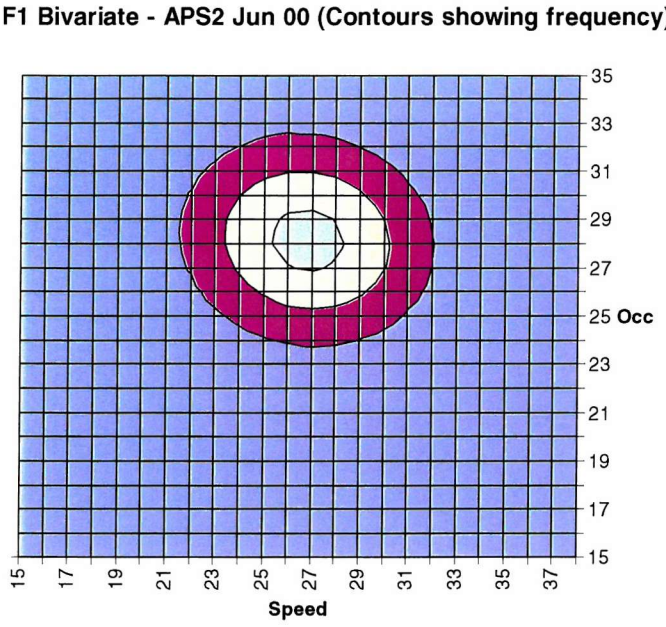


Figure 47 – Contour maps for the distribution of f_1 for APS1 and APS2 using June 2000 data, the shape of the contour indicates little correlation between the parameters in this set

Again the additional data for APS2 has brought a slight change in the overall shape. In both diagrams it is clear that there is very little correlation between the two variables in this data set. The correlation value of -0.08 confirms this to be the case. In fact, as the value of the correlation coefficient tends to zero the bivariate normal distributions tend towards the product of the two normal distributions. Based on these observations, it is possible to conclude that the speed-density relationship may not necessarily govern traffic behaviour during the transition between free-flow and congestion.

6.3.3 Using the posterior as a new prior

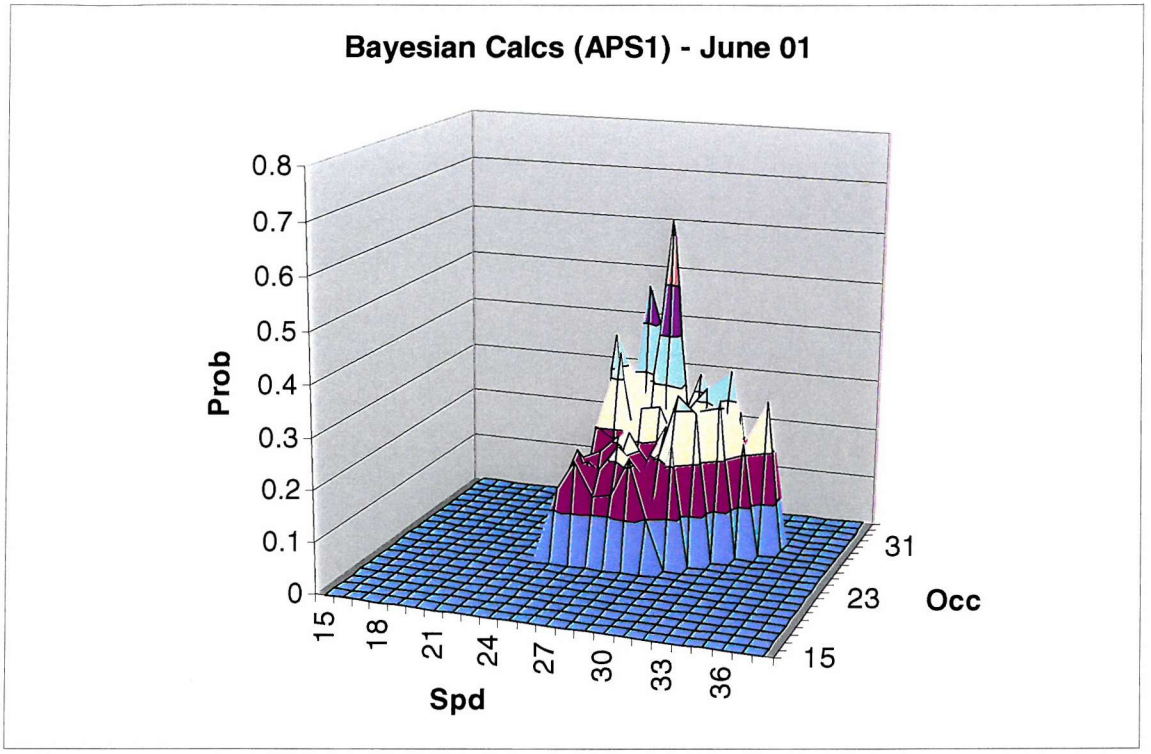
The method described in Section 6.2 was applied to traffic data from two different months (using a uniform value for the prior as before). The process was then repeated but this time using a refined prior, namely the posterior matrix from June 2000. This refinement represents a more accurate statement of belief about the distribution of the random 'flow-breakdown' variable for (θ). Figure 48 gives a comparison of posterior distributions obtained for June 2001 using the simple and refined prior. Figure 49 gives similar information but for September 2001. Additional plots for likelihood are found in Appendix E.

The application of a refined posterior leads to an interesting redefinition of the probability matrix. The peak region (corresponding to the highest likelihood of seed-points occurring) is **amplified**. This is an important result, particularly in terms of being able to use the matrix in traffic applications.

One of the immediate applications of the probability map derived from the posterior distribution is the early detection of shockwaves. A refined posterior distribution provides a control room operator with greater confidence in the results because of the more distinct peak values. If the traffic application gives too many false positives (high probabilities but no subsequent shockwave) then it will be discredited. On the other hand, if the traffic application gives too many false negatives (low probabilities but shockwaves subsequently occur) then it will be considered a redundant application. Incident detection algorithms are capable of detecting the shockwave once it has begun to propagate, so the particular niche occupied by this probability map is that of early detection.

Finding the balance between the false positive and the false negative is achieved by:

- Using a refined posterior matrix and updating it regularly;
- Creating template matrices tailored to the specific characteristics of the motorway section under consideration;
- Carefully considering the thresholds presented to an operator (these will be an interpretation of the probability values offered).



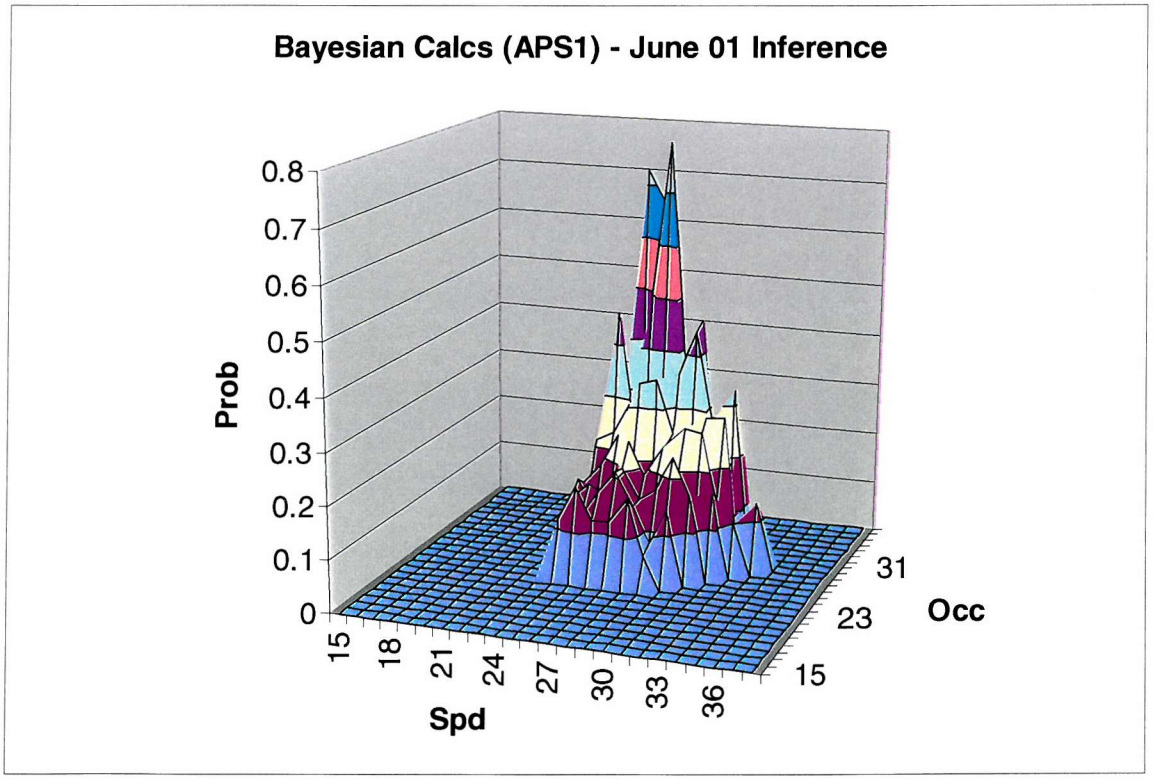
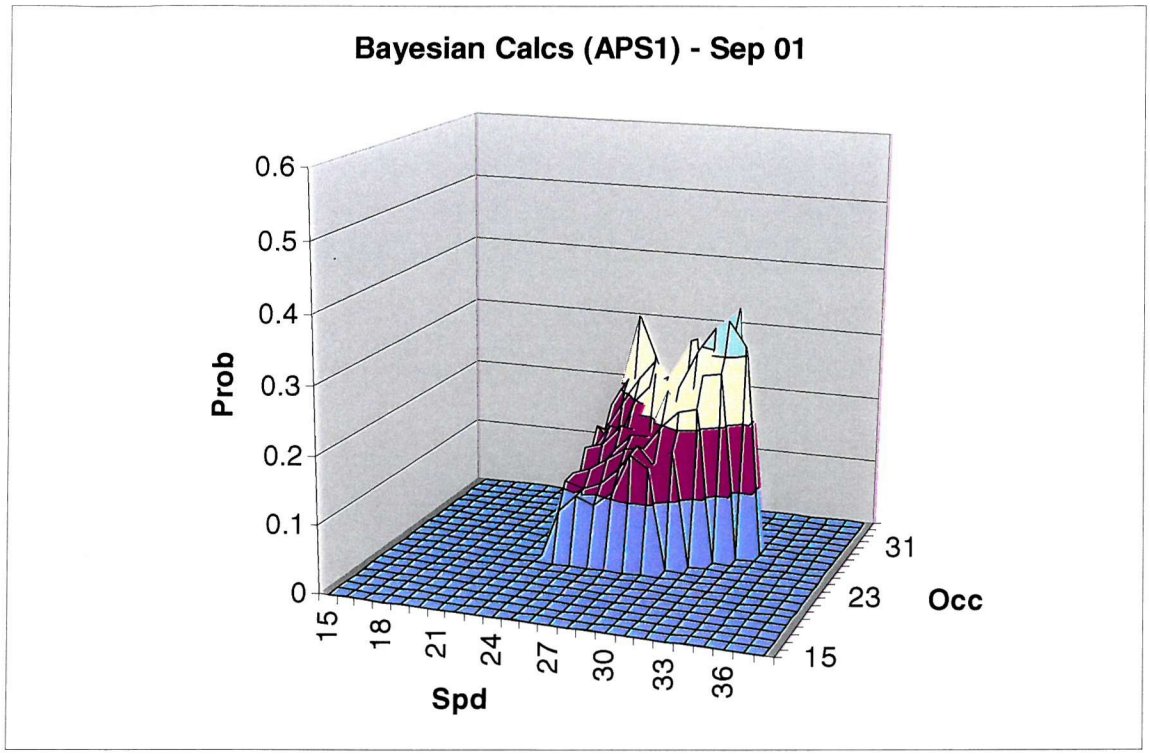


Figure 48 – Posterior for June 2001 derived using a simple prior and a refined prior (based on the posterior for June 2000)



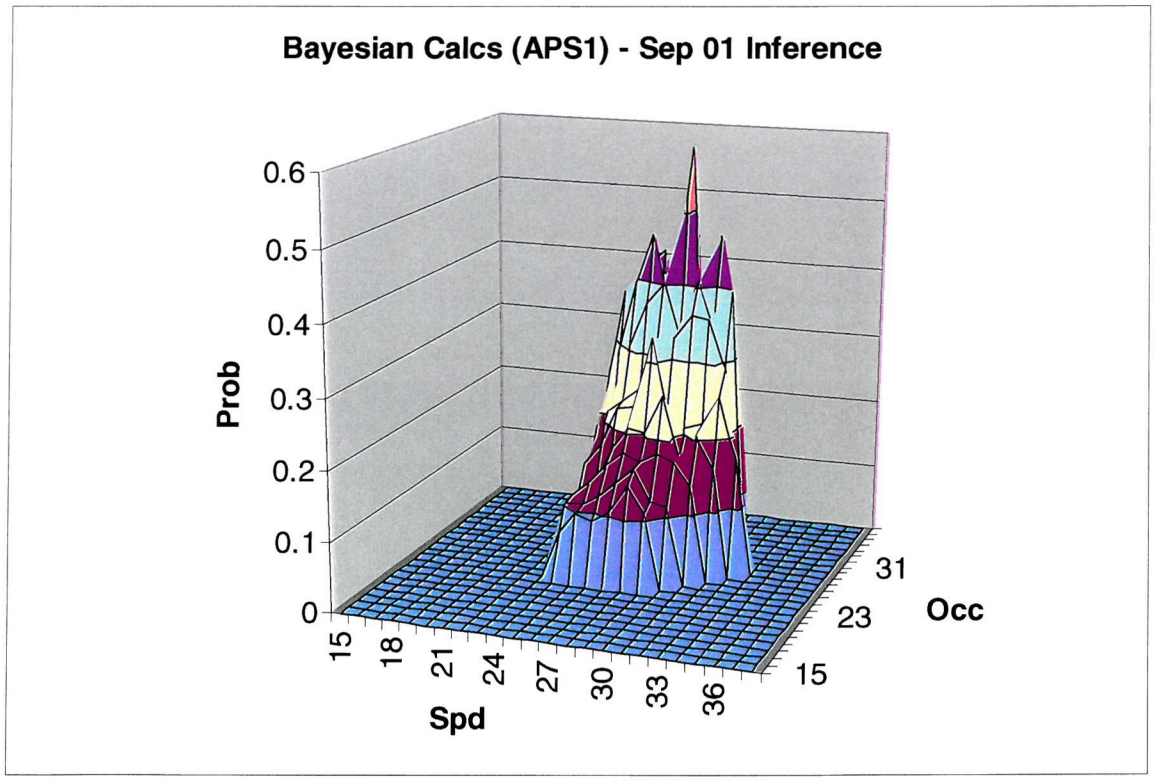


Figure 49 – Posterior for September 2001 derived using a simple prior and a refined prior (based on the posterior for June 2000)

6.3.4 Relationship with upstream data

Earlier in the thesis a hypothesis was made that there would be a relationship between traffic conditions identified upstream of a particular location and the subsequent occurrence of flow-breakdown (shockwaves) at that location. Recall from Section 5.3 that each point identified by the software was marked out with a grid, representing three consecutive loops by three consecutive minutes. The centre point (P0) held the data satisfying the conditions (Rules 1–5). The above hypothesis translates into a possible relationship between P0 and P6.

Figure 50 illustrates this in a most efficient manner. If there was any possibility of a relationship between the speed/occupancy pairs collected at P6 (corresponding to successful identification of a seed-point using the rules at P0) then we would expect to see some clustering of the 'pink series'. However, the variance in speed and occupancy is large (an order of magnitude greater than consecutive minutes for the same loop). It must therefore be concluded that seed-points identified at P0 are independent of any variation in traffic characteristics (speed, occupancy) occurring at P6.

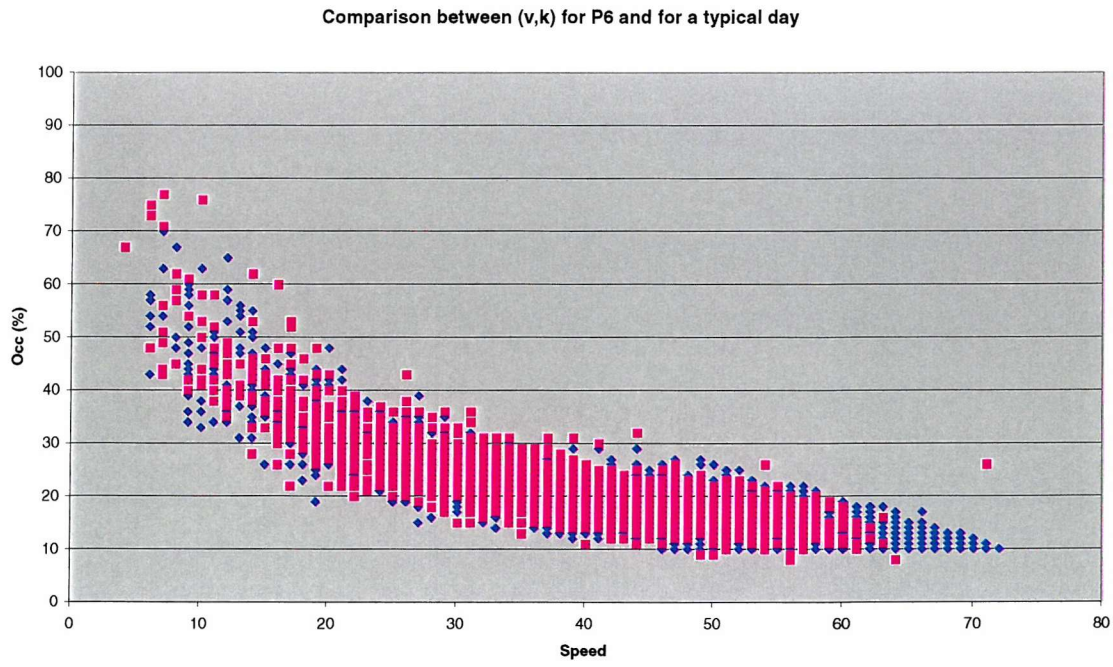


Figure 50 – Speed/occupancy pairs from P6 overlaid on speed/occupancy data for the whole day (MIDAS 1-Minute Average)

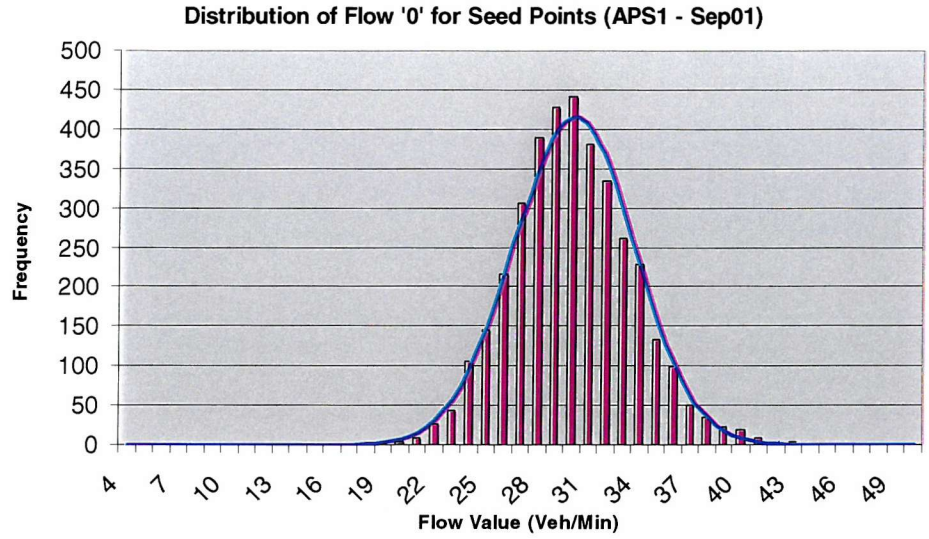
Data resolution has been constrained by the spacing of loops along the study location (around 500m between consecutive loops). Although this is currently one of the best sites on the HA Network for detector coverage, there are plans within the new Active Traffic Management Pilot (M42) to install loops at a spacing of around 100m. This brings in a whole new set of considerations (especially the need to ensure that vehicles are not double-counted within the same minute), and changes might need to be applied to the averaging process. However, given this new resolution, it may be possible to revisit the predictive hypothesis suggested in this thesis, since information which is lost at a spacing of 500m might surface at a spacing of 100m.

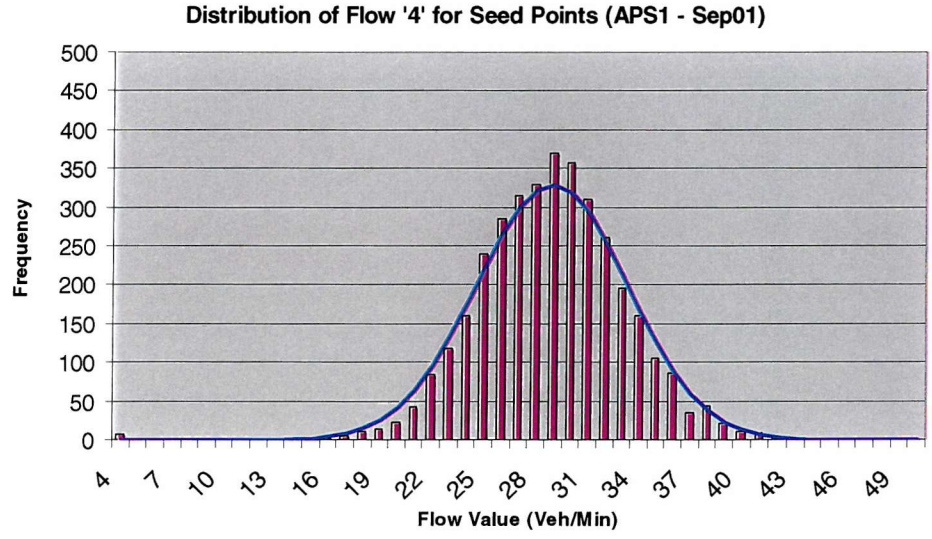
The number of pairs qualifying as seed-points under APS1 for September 2001 was 3,699. Looking at the data it is interesting to note that the flows are normally distributed at the point of classification (P0), and follow a similar pattern at the upstream detector (P4, P6), see Figure 51. Flows associated with the commencement of flow-breakdown are in the range of around 25–35 veh/min (1500–2100 veh/hr) with a concentration around 1750 veh/hr.

Earlier in this thesis, flow was discarded as a primary detection variable. It is common to look for the threshold of flow-breakdown at the extreme point of the speed-flow curve (maximum sustainable throughput of a motorway section), where speeds are in the region of 50mph and flows in the region of 2500+ veh/hr/lane. It is therefore interesting to see that flows associated with seed-points are considerably lower. This is further illustrated in Figure 52, with a typical speed-flow curve for a single detector in the morning peak. For this particular day the transition takes place at 06:42 at a flow of around 1700 veh/hr. Although a peak was reached in the previous minutes, the direction of flow-breakdown is first an oscillation along the upper curve and then an *arcing* to the lower region at 06:42.

Until congestion occurs, attempts by the driver to achieve desired speed and sustain minimum headway exert a direct and repeated influence on the stability of traffic flow. These observations confirm that variation in flow is not a causal variable, but sufficient flows are required to sustain propagation of shockwaves once they have been triggered.

UNIVERSITY OF SOUTHAMPTON LIBRARY





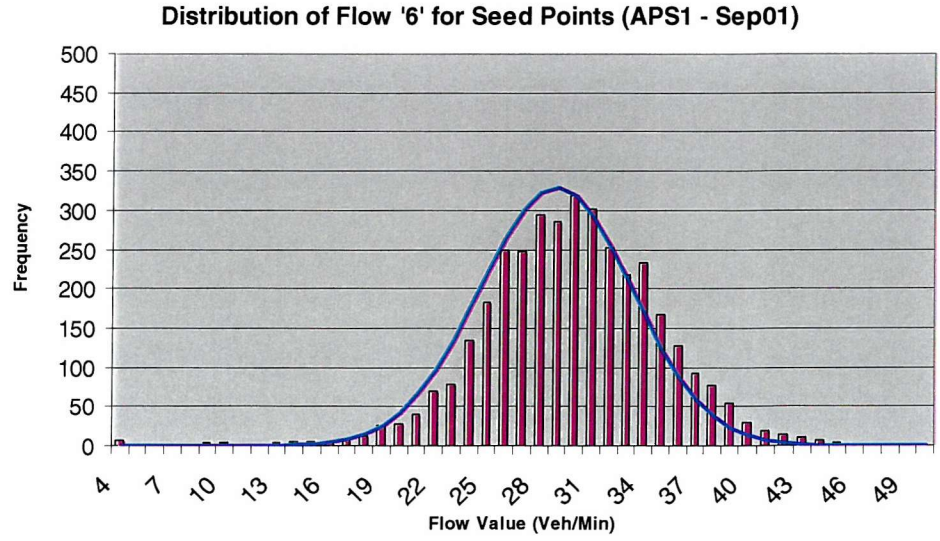


Figure 51 – Flow distribution associated with seed-points from Sep 2001

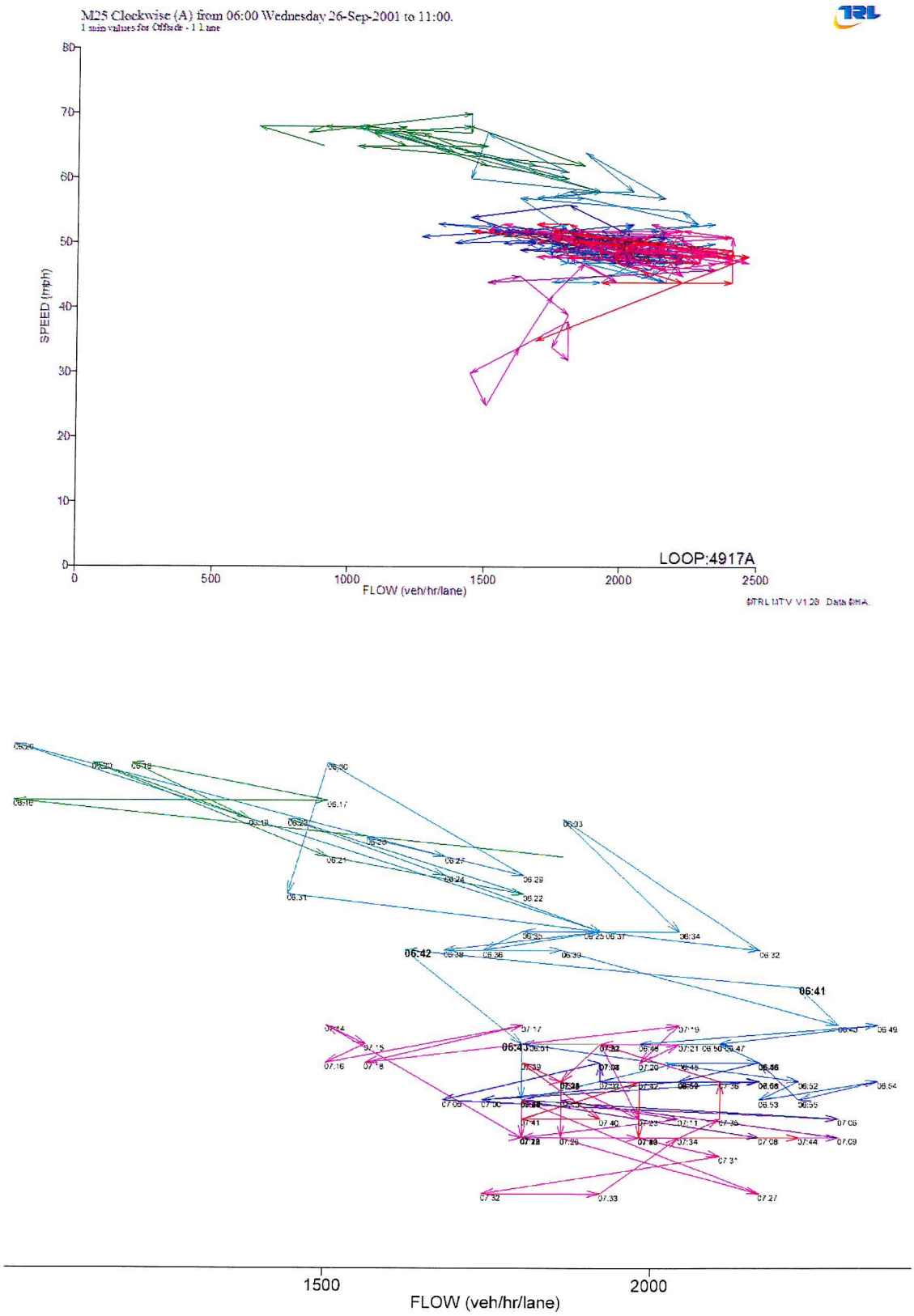


Figure 52 – Example speed flow curve with amplification on transitional region between 06:30 and 07:45

6.3.5 Combining probability map with incident detection

Figure 53 shows the composite output from the probability matrix and the incident detection algorithm (in this case California) compared with the equivalent occupancy detector data for the same period. The dark pink squares are from the incident detection algorithm, whereas the more scattered varying shades of pink are obtained from the probability matrix.

The probability matrix fills in the regions missed by the incident detection algorithm although the latter detects the body of the shockwave as it has been designed to do.

In terms of control room applications a *combination* of these two methods would be very powerful, giving the operator an indication of when flow-breakdown was probable and then confirming and tracking the subsequent shockwave.

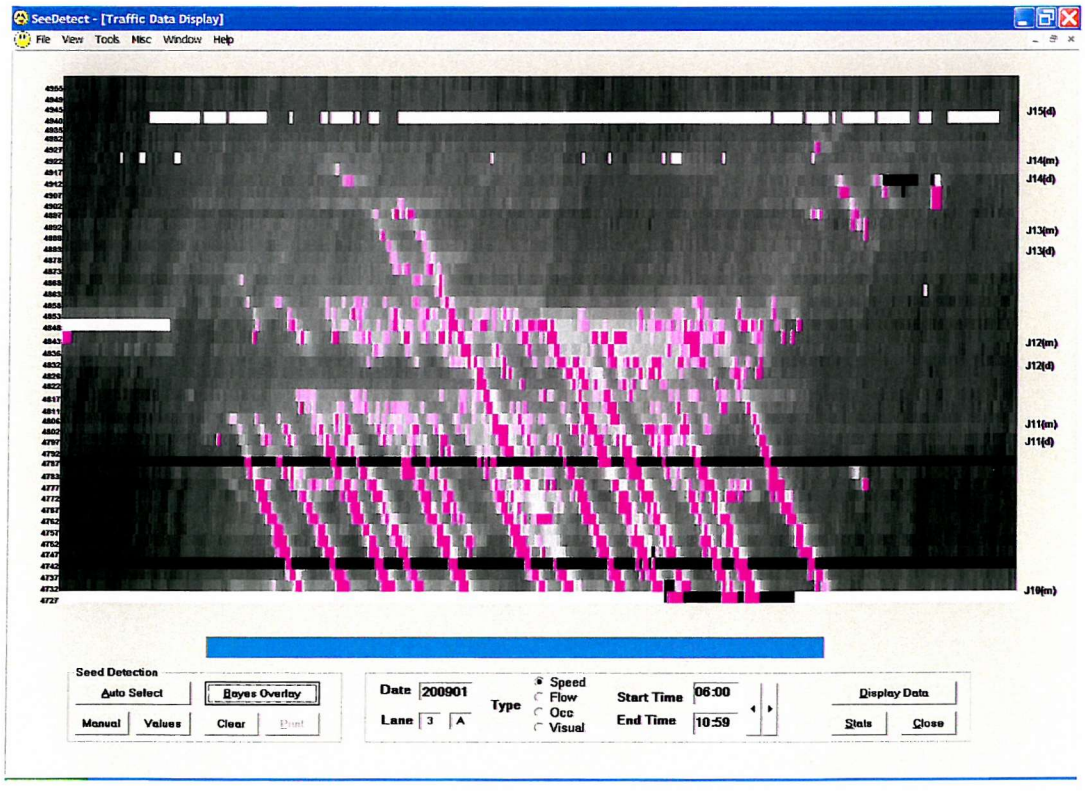


Figure 53 – Combination of probability map and output from incident detection algorithm

6.4 SUMMARY OF RESULTS

The basic methodology for creating a probability map for seed-points has been illustrated in three steps using data from June 2000. The following observations have been made.

- The reason for choosing parameter sets was simply to provide a filter within which to operate the rules. Looking at two parameter sets (APS1 and APS2) gave some understanding about the sensitivity behind the choices made on the basis of the speed-headway investigation. Results obtained are sufficient for demonstrating the methodology in this study, but there is scope for conducting an exercise in optimisation and exploring variation in all the parameters.
- The bivariate normal distribution has been examined for its benefit in allowing rapid expansion of the data set (using only the statistical parameters). It is not necessary if there is good quality data available and if the objective is to derive a series of monthly templates. But it may have an application where data is poor or when the objective is to build a single historical archive.
- By using a posterior matrix as the refined prior for a new data set the resulting output has a more clearly accentuated peak. When it comes to application this is an important result.
- No immediate relationship was identified between the occurrence of flow-breakdown and traffic data at the upstream detector. The introduction of detectors spaced at 100m (instead of 500m as used in this study) provides opportunity to explore such a relationship in the future. However, the probability map does act as an early warning system at the same detector. It should be possible to increase the warning time by refining the algorithm and tuning the thresholds.
- Again with application in mind, the combination of incident detection and the probability matrix provides coverage for both seed-points and shockwave patterns within the traffic data.

7 CONCLUSIONS AND FUTURE RESEARCH

7.1 SUMMARY OF KEY RESULTS

7.1.1 Empirical survey of seed-points

The study began by introducing the problem of shockwaves in motorway traffic. A comprehensive survey of the literature and current knowledge about shockwaves was presented in Chapter 2.

An empirical investigation into the causes of congestion was conducted on the M25 J10–15 in order to gain a better understanding of the mechanisms behind shockwave origins. Numerically, the investigation found that *recurrent congestion* accounted for around 78% of congestion delay, accidents for 20% and breakdowns for 1%. A range of miscellaneous causes accounted for the remaining 1% between them. Closer analysis of recurrent congestion showed that flow-breakdown at merges, diverges and between junctions were each responsible for about one third of the total for recurrent congestion. In other words, the problem is distributed evenly across the three main contributory factors. Details of this investigation were presented in Chapter 3.

7.1.2 Developing the probabilistic model

On the basis of this understanding, a conceptual framework was developed for modelling seed-points, comprising two key elements. The first (known as the Mechanistic Approach) looked at how the mechanisms proposed within car-following theory can describe the origin of shockwaves (rather than their propagation). The second (known as the Probabilistic Approach) looked at whether it was possible to make an inference about the existence of a seed-point at a particular time and location, given certain traffic parameters. The probabilistic method was taken forward through application of Bayesian Inference to this particular phenomenon.

Instead of dividing traffic in two states and looking at 'congestion', this study considered traffic in three states (free, synchronised and congested) as increasingly recommended in the literature. 'Flow-breakdown' was described in terms of a binary random variable, but with the addition of an important

assumption. A value θ=0 was used for free-flowing *and* congested traffic, so the body of the shockwave was excluded. A value of θ=1 was used for synchronised flow (the intermediary and fluctuating state prior to congestion). This differs from the conventional approach of treating the flow-breakdown variable in terms of vehicle arrival rates. Boundaries between the various states (and therefore the definition of θ) were dependent on the detection method employed.

7.1.3 Seed point detection

Testing the probability model required the accurate identification and classification of shockwave origin points. Since most of the existing detection methods focus on the shockwave body, a new approach was required for seed-points. Special software called SeeDetect© was developed by the author in order to examine the characteristics of a seed-point. Manual selection of seed-points gave some understanding of patterns involved. A detection algorithm based on this knowledge (in the form of five rules and a series of parameters) has been presented as part of this study.

Investigation of the MIDAS data archives showed that the flow variable (veh/min) was not suitable for the identification of shockwave origins, since there was no significant variation in flow rate until the shockwave itself was more pronounced.

The other key variables available from the MIDAS detectors are average speed and occupancy (percentage of a one-minute interval for which a detector is physically covered). Occupancy is a proxy for vehicle headway, so a detailed examination of speed-headway relationships was conducted using individual vehicle data. With this new understanding of the mechanisms involved, a range of parameters was chosen for the detection algorithm in order to collect the appropriate traffic information for the probability model.

7.1.4 The probability map

Having obtained all the component parts for the Bayesian Model, a posterior distribution was produced describing the likelihood of a seed-point given a particular combination of speed and occupancy. This matrix was imported by

the SeeDetect© software and used to produce 'Probability Maps' for the study site (M25 J10–15). Clearly the results depended heavily on the assumptions behind the classification algorithm, the thresholds limiting the data sets, and the quality of data available. This thesis has provided an adequate demonstration of the novel methodology, and the work can be taken forward to useful implementation by means of a calibration exercise.

One key characteristic of Bayesian Inference is the need to think about the nature of the prior and then allow the data (likelihood) to reshape that belief into a posterior. This characteristic has been exploited by using the posterior from one data set as the refined prior for another data set. When a constant value for the prior was assumed, the results reflected that original belief but reshaped them into a more bell-shaped function with clear boundaries. When this refined shape was assumed as the prior for new data, the results reflected that new belief and offered small refinements specific to that data set. Monthly and seasonal variations in traffic data are common knowledge and should be taken into account. However, the general shape of the probability map is consistent – which is encouraging.

7.1.5 General summary

The method is summarised as follows:

- (Step 1) Detecting seed-points within a defined search space and sorting the data;
- (Step 2) Calculating key components for probabilistic equation, and deriving the subsequent posterior matrix;
- (Step 3) Creating a probability map using real detector data and evaluating the predictions against occurrence of shockwaves.

Initially the study focussed on deriving a relationship between traffic conditions upstream of the seed-point to the occurrence of flow-breakdown from that point. However, it soon became apparent that no such relationship could be identified with any confidence (given the data resolution), and the benefits of the probabilistic method were emphasised in terms of early detection rather than prediction.

The initial focus of this study was the development of a statistical model to describe the relationship between traffic conditions detected upstream of a particular location, and the subsequent occurrence of flow-breakdown (in the form of shockwaves) at the location. However, it soon became clear that due to the resolution of the data (spacing of detectors) it would not be possible to infer anything about the existence of such a relationship, and the focus of the study moved from prediction to early detection. In terms of this new objective, the Bayesian Model presented in this thesis can be considered successful.

Early detection and the confirmation of seed conditions based on historical information are a useful contribution to the whole sphere of motorway control systems.

7.2 DIRECTION OF FURTHER RESEARCH

7.2.1 Refinements to current approach

This thesis presents a novel application of inferential statistics and traffic theory. The resultant probabilistic model for the origin of motorway shockwaves can be applied to *an individual detector* and used to build up a probability map by evaluating the minute-by-minute data as it appears. This evaluation takes the form of a probability or likelihood that a seed-point has occurred (the exact definition of which will depend on the assumptions used in developing the posterior distribution from which the model operates). Research conducted during this study has also helped to tackle the question of how variable speed-limits might be used to minimise the effect of shockwaves. Having demonstrated the method it is now important to review where improvements can be made.

Clearly the results are highly dependent on the accuracy of classification so further work on the optimisation of thresholds is warranted. The five rules which make up the seed detection algorithm could also be refined further. An example would be the inclusion of a feedback mechanism based on subsequent flows to indicate that a shockwave had in fact taken place. This might be embedded in the algorithm or built into a parallel incident detection system. Successful predictions could then be weighted and subtle alterations

could then be made over time to the posterior function which forms the probability map. Such an arrangement would better reflect conditions in a control room where the benefit of hindsight is not available and predictions need to be made on an unfolding minute-by-minute basis.

A probability matrix was derived from data for June 2000 and then used as a refined prior for data from different months in order to make predictions. The next step here would be to develop some **probability matrix templates** for different seasons, junction types, or even motorways. For example, the traffic behaviour on the M42 ATM Site is quite different from the behaviour on the M25 Controlled Motorway. Any future traffic control application could then select appropriate templates according to some logic which categorises the conditions.

The site chosen for this study was subject to speed control for certain periods during the day. Assuming that active traffic management does produce lower speeds due to compliance with variable speed limits, a smoother flow profile and a reduction in the overall occurrence of flow breakdown, then there would be some bearing on application of the proposed model. Much of this impact would be eliminated through the development of templates identified in the previous paragraph. Much of the influence from variable speed limits would come from the 40mph settings (closer to the speed range associated with the seed-points), rather than the 50mph or 60mph settings. Where a site was not governed by speed control, it is reasonable to expect that (with all other things being equal) the sample size for f1 (seed-points) would be greater. It would also be reasonable to expect a greater spread in the corresponding distributions reflecting greater variation in the traffic parameters generally.

SeeDetect© is software developed to assist in the analysis of detector data but there may be scope for integrating such functionality within a 'real-time' version of the MTV Software. This could then be provided to control-centre operators so that they can maintain and control the network. This work would involve taking the probability template from previous research and combining it with historical information.

7.2.2 Other avenues of shockwave research

Work on shockwave behaviour is currently focussed on origin and propagation. However, it would be profitable to investigate the *frequency* of shockwave occurrence, perhaps by using time-series analysis, systems theory, or better still using queuing theory. This latter approach might consider the behaviour of traffic entering a shockwave area as analogous to the behaviour of traffic at traffic lights. Each origin is thought of as a set of traffic lights and the time-varying frequency of the formation of shockwaves at each origin is a time-varying green cycle time.

In parallel to this theory, there might also be improvements to the physical description of how traffic arrives at and leaves each shockwave. Kerner (1997) investigated this behaviour with regards to simple queues on motorways, and his work would provide a good starting point. Whitham (1999) gave an excellent review of linear and non-linear wave theory applications in traffic, giving particular focus to the traffic light problem and assumptions to improve the realism of kinematic wave models. Rees *et al.*, (2000) observed that following the opening of the M4 Bus Lane there was a change of traffic behaviour between J3 and J2 from ordinary congestion to shockwave behaviour. Conducting an investigation into shockwave frequency may shed light on the mechanism behind this change.

Another area of study would be to revisit the mechanistic approach to seed-point modelling (see Section 4.2). This thesis has proposed improvements to the speed-headway relationship adopted in car-following models (treating each lane differently). Implementation of such improvements could be accompanied by further analysis of the stability of such algorithms, and whether indeed shockwaves can be naturally induced. Although the work was initially proposed with the microsimulation model SISTM in mind, other models may well be more suitable (particularly if they can already handle shockwave propagation).

7.3 POSSIBLE APPLICATIONS

7.3.1 Active Traffic Management (ATM)

In terms of dealing with the question of recurrent congestion, the three main options are to

- Reduce the overall demand from road users;
- Build new capacity to match the demand for traffic;
- Make more efficient use of the existing network.

Current government policy of transport puts greatest emphasis on making better use of the existing network to absorb rising demand. Figure 54 is a virtual representation of the Active Traffic Management (ATM) environment. ATM is a proactive and innovative approach to tackling congestion in a flexible, coordinated, and efficient manner. Alongside the National Traffic Control Centre project, ATM forms a fundamental part of the Highways Agency delivery programme over the coming few years. Some of the proposed traffic management techniques will be familiar to drivers, while others will be a new experience in terms of changes to the driving environment. However, ATM is a versatile and integrated solution, aimed at tackling specific problems identified at the chosen location.

The ATM Pilot is to be located on the M42 J3A to J7. This section of motorway forms an important part of the regional network, with arterial links into Birmingham and access to major attractions such as the National Exhibition Centre (NEC) and Birmingham Airport.

Looking at Figure 54 again, the basic idea of control is variable speed-limits and message signs used in combination to communicate appropriate instructions to drivers based on traffic conditions. In 2004 the primary operational regimes for implementation are the familiar variable speed-limits (for three lanes) and the more novel 'actively managed hard-shoulder' (to be operated during congestion or incidents).

Figure 54 – A virtual representation of the Active Traffic Management environment (courtesy of Mouchel Consulting Ltd)

It is currently planned that the control room operators will make decisions about when the hard-shoulder should be used. Judgement would be required in order to optimise between opening too early (thereby prematurely applying a speed-limit), and opening too late (once flow-breakdown has set in). Following refinements suggested in the previous section, the findings of this thesis may have some application in assisting operators with this decision process.

7.3.2 Dynamic vehicle control

This study began by presenting the context for the requirement to predict shockwave occurrence (see Section 3.2), namely the feasibility of curtailing the propagation of shockwaves through dynamic traffic control using existing infrastructure such as Controlled Motorways. This can be taken one step further by considering the possibility of external vehicle speed control, in-vehicle technology, and intelligent highway infrastructure.

Car manufacturers are striving to improve their vehicles and to gain a competitive advantage. The most significant such development for motorway operation is Adaptive Cruise Control (ACC) but a serious question remains about the acceptability of shorter headway between vehicles, and the extent to which the driver can be depended on to make a safe selection. If the safety of such systems could be quantified for time headway below two seconds (see Section 5.4.2) this would open up the potential for effective control and the optimisation of safety versus capacity.

Marsden, McDonald and Brackstone (2001) simulated potential impacts of ACC on motorway driving and contrasted the findings with drive-cycles from real vehicles. They concluded that driving with an ACC system can provide considerable reductions in the variation of acceleration (compared with unassisted driving). This has obvious implications for fuel efficiency, vehicle emissions, driver workload, and even safety. The investigation also concluded that there would be no short-term improvements in traffic efficiency during the peak periods. Field trials of ACC systems gave evidence that drivers are slow to re-engage the system once manual intervention has been applied and perhaps preferred to remain fully in control during busy periods. This behavioural pattern may well change with familiarity, confidence, and increased market penetration.

Ideally vehicle speeds and headways should be adjusted dynamically to road and traffic conditions. Dynamic control requires sensors to determine the current conditions, algorithms to process the information and a communications system to inform vehicles. Information could be transmitted by a Dedicated Short Range Communication system, as used in the Road Traffic Advisor project and in electronic tolling systems, by mobile telephone technology or radio broadcast. However, the major need is to determine how best to use the information.

Research is needed into how accurate the information should be, how frequently, in time, it needs to be updated and the separation of the sensors,. A deeper understanding of the liability issues of providing dynamic information and control is also required. It may be argued that drivers are entitled to rely

on the information if dynamic speed-limits or headways are supplied to vehicles.

Furthermore, understanding the frequency of shockwaves and the mechanism by which traffic moves between shockwaves, may enable suitable control strategies to be applied to smooth out the traffic, increase throughput and reduce the occurrence of these events.

8 REFERENCES

Abou-Rahme, N., Beale, S., Harbord, B. & Hardman, E. (2000) Monitoring and Modelling of Controlled Motorways, *RTIC 2000*, Publication 472, IEE London, 84–90.

Abou-Rahme, N., Hardman, E., Robson, B. & Still, P. (1996) *The ability of SISTM to replicate speed/flow behaviour on the M25*, Unpublished Report PR/TT/089/96, Transport Research Laboratory, Crowthorne, UK.

Abou-Rahme, N. & White, J. (1999) Replicating shockwaves within a classical car following model, *Annual Research Review*, 1999, Transport Research Laboratory, Crowthorne, UK, 72–82.

Abou-Rahme, N., White, J. & Stewart, R., (2002) Journey Time Estimation Using MIDAS Loop Data, *RTIC 2002*, Publication 486, IEE London, 177–181.

Addison, P. & Low, D. J. (1994) Order and chaos in the dynamics of vehicle platoons, *Traffic Engineering and Control*, **37**(718), 456–459.

Ashton, W. D. (1966) *The theory of road traffic flow*, London: Methuen.

Bando, H., Hasebe, K., Nakayama, A., Shibata, A. & Sugiyama, Y. (1995) A dynamical model of traffic congestion and numerical simulation, *Physics Review E*, **51**(2), 1035–1042.

Banks, J. H. (1991) The two capacity phenomenon: some theoretical issues, *Transportation Research Record*, **1320**, Transportation Research Board, Washington, USA, 234–241.

Banks, J. H. (1995) Another look at A Priori relationships among traffic flow characteristics, *Transportation Research Record*, **1510**, Transportation Research Board, Washington, USA, 1–10.

Bayes, T. (1763) An essay towards solving a problem in the doctrine of chances, *Philosophical Transactions of the Royal Society of London*, **53**(1763), 370–418 [reprinted in *Biometrika*, **45**, 1958, 292–315].

Benehokal, R. F. & Treiterer, J. (1988) CARSIM: Car-following model for simulation of traffic in normal and stop-and-go conditions, *Transportation Research Record*, **1194**, 99–111.

Brackstone, M. & McDonald, M. (2003) Driver behaviour and traffic modelling: are we looking at the right issues? *Proc. of the IEEE Intelligent Vehicles Symposium 2003, Columbus, Ohio, U.S.A.*, June 9-11. CD-ROM, IEEE.

Brackstone, M., Sultan, B. & McDonald, M. (2002) Motorway driver behaviour: studies in car following, Transportation Research Part F, 0**5F** (1), 329-344.

Buckley, D. J. (1968) A semi-Poisson model of traffic flow, *Transportation Science*, **2**(2), 107–133.

Busch, F. & Fellendorf, M. (1990) Automatic incident detection on motorways, *Traffic Engineering and Control*, April 1994, UK.

Chandler, R. E., Herman, R. & Montroll, E. W. (1953) Traffic dynamics: studies in car-following, *Operational Research*, **6**, 165–184.

Chang, G. L., Wu, J. & Cohen, S. L. (1994) Integrated real-time ramp metering model for non-recurrent congestion: Framework and preliminary results, *Transportation Research Record*, **1446**, Transportation Research Board, Washington, USA, 56–65.

Charlesworth, G. (1950) Methods of making traffic surveys especially 'Before and After' studies, *Journal of the Institution of Highway Engineers*, **1**, 42.

Chin, H. C. (1996) Analysis of highway bottlenecks, *Proceedings of the Institution of Civil Engineers – Transport*, **117**, May, 136–142.

Chin, H. C. & May, A. D. (1991) Examination of the speed-flow relationship at the Caldecott Tunnel, *Transportation Research Record*, **1320**, Transportation Research Board, Washington, USA, 75–90.

Dixon, C., Harbord, B. & Abou-Rahme, N. (2002) *Speed control and incident detection on the M25 Controlled Motorway (Summary of results, 1995–2002)*, PR/T/095/02, Transport Research Laboratory, Crowthorne, UK.

Dudek, C. L., McCasland, W. R. & Burns, E. N. (1988) Promotional issues related to accident investigation sites in urban freeway corridors, *Transportation Research Record*, **1173**, Transportation Research Board, Washington, USA, 1–10.

Edie, L. C. (1961) Car following and steady state theory for non-congested traffic, *Operations Research*, **9**, 66–76.

Edie, L. C. (1963) Operations research in tunnel traffic control, *Journal of the Engineering Mechanics Division, Proceedings of the American Society of Civil Engineers*, **89**, Part 1, December, 15–28.

Elefteriadou, L. (1997) Freeway-merging operations, a probabilistic approach, *8th International Federation of Automatic Control (IFAC)*, Technical University of Crete, Greece, June 1997, 1351–1356.

Elefteriadou, L., Roess, R. P. & McShane, W. R. (1995) Probabilistic nature of breakdown at freeway merge junctions, *Transport Research Record*, **1484**, Transportation Research Board, Washington, USA, 80–89.

Fazio, J., Michaels, R. M., Reilly, W. R., Schoen, J. & Poulis, A. (1990) Behavioural model of freeway exiting, *Transportation Research Record*, **1281**, Transportation Research Board, Washington, USA, 16–27.

Gall, A. L. & Hall, F. L. (1989) Distinguishing between incident congestion and recurrent congestion: a proposed logic, *Transportation Research Record*, **1232**, Transportation Research Board, Washington, USA, 1–8.

Gazis, D. C., Herman, R. & Potts, R. B. (1959) Car-following theory of steady state traffic flow, *Operations Research*, **7**, 499–505.

Gazis, D. C., Herman, R. & Rothery, R. W. (1961) Non-linear follow the leader models of traffic flow, *Operations Research,* **9**, 545–567.

Gipps, P. G. (1981) A behavioural car-following model for computer simulation, *Transportation Research Part B,* **15B**, Pergamon Press, Oxford, 105–111.

Greenberg, H. (1959) An analysis of traffic flow, *Operations Research*, **7**, 79–85.

Haight, F. A. (1958) Towards a unified theory of road traffic, *Operations Research*, **6**, 813-826.

Haight, F. A., Whisler, B. F. & Mosher, W. W. (1961) New statistical method for describing highway distribution of cars, *Proceedings of the Highways Research Board, Washington,* **40**, 557-564

Hardman, E. & Taylor, M. (1992) SISTM: A computer simulation of traffic on motorways, Working Paper WP/TO/94 (unpublished), Transport Research Laboratory, Crowthorne, UK.

Helbing, D. (1997) Traffic data and their implication for consistent traffic flow modelling, *8th International Federation of Automatic Control (IFAC)*, Technical University of Crete, Greece, June 1997, 809–814.

Herman, R., Montroll, E., Potts, R. B. & Rothery, R. W. (1959) Traffic dynamics: analysis of stability in car-following, *Operations Research*, **1**, 86–106.

Holland, E. (1998) A generalised stability criterion for motorway traffic, *Transportation Research Part B*, **32B**(2), 141–154.

Hounsell, N. B., Barnard, S. R., McDonald, M. & Owens, D. (1992) *An investigation of flow breakdown and merge capacity on motorways,* Contractor Report **CR338**, Transport Research Laboratory, Crowthorne, UK.

Hounsell, N. B., Barnard, S. R. & McDonald, M. (1994) Capacity and flow breakdown on UK motorways, *Proceedings of the Second International Symposium on Highway Capacity*, (ed. Akçelik, R.), **1**, Sydney 1994, Australian Road Research Board, 277–294.

Hsu, P. & Banks, J. H. (1993) Effects of location on congested-regime flow-concentration relationships for freeways, *Transportation Research Record*, **1398**, Transportation Research Board, Washington, USA, 17–23.

Hunt, J. G. & Yousif, S. Y. (1994) Traffic capacity at motorway roadworks – effects of layout, incidents and driver behaviour, *Proceedings of the 2nd International Symposium on Highway Capacity*, Sydney, Australia **1**, 295–314.

Jarrett, D.F. & Zhang, X. (1997) Stability analysis of the classical car-following model, *Transportation Research Part B*, **31B** (6), 441–462.

Johnson, A. N. (1928) Maryland aerial survey of highway traffic between Washington and Baltimore, *Highway Research Board*, Proceedings, **8**, 106–115.

Kalman, R. (1960) A new approach to linear filtering and prediction problem, *Journal of Basic Engineering*, **82D**, ASME, 1960, 34–35.

Kerner, B. (1997) Experimental characteristics of traffic flow for evaluation of traffic modelling, *8th International Federation of Automatic Control (IFAC)*, Technical University of Crete, Greece, June 1997, 793–798.

Kerner, B. (1999) The physics of traffic, *Physics World*, August 1999, 25–30.

Kühne, R. (1991) Traffic patterns in unstable traffic flow on freeways, Highway Capacity and Level of Service, Brannolte (ed.), *Proceedings of the International Symposium on Highway Capacity*, Karlsruhe, 24–27 July 1991, Balkema, Rotterdam, 211–224.

Kulkarni, R. G., Stough, R. R. & Haynes, K. E. (1996) Spin glass and the interactions of congestion and emissions: An exploratory st*ep*, *Transportation Research Part C*, **4**(6), 407–424.

Leutzbach, W. (1988) *Introduction to the Theory of Traffic Flow*, London: Springer-Verlag.

Lighthill, M. J. & Whitham, G. B. (1955) On Kinematic Waves: II a theory of traffic flow on long crowded roads, in *An Introduction to Traffic Flow Theory* (eds. Gerlough, D. L. & Capelle, D. G.), Highway Research Board Special Report, **79**, Washington 1964.

Luttinen, R. T. (1994) Identification and estimation of headway distributions, *Proceedings of the second international symposium on highway capacity*, (ed. Akçelik, R.), Volume 2, Sydney 1994, Australian Road Research Board, 427–446.

Marsden, G., McDonald, M., & Brackstone, M. (2001) Towards adaptive cruise control, *Transportation Research Part C*, **09C**, 33–51.

Mason, A. & Woods, A. (1997) A car following model of multi-species systems of road traffic, *Physics Revue E*, **55**(3A), 2203–2214.

May, A. D. (1990) *Traffic Flow Fundamentals*, London: Prentice Hall International.

May, A. D. (1994) TRB Distinguished Lecture, *Transportation Research Record*, **1457**, Part 1, Transportation Research Board, Washington, USA.

Michaels, R. M. & Fazio, J. (1989) A driver behaviour model of merging, *Transportation Research Record*, **1213**, Transportation Research Board, Washington, USA, 4–10.

Newell, G. F. (1961) Non-linear effects in the dynamics of car-following, *Operations Research*, **9**, 209–229.

Nihan, N. L. (1993) Forecasting freeway and ramp data for improved real-time control and data analysis, Volume 1 – Summary Report, *NTIS* (PB94–106523).

Ostrom, B., Leiman, L. & May, A. D. (1993) Suggested procedures for analysing freeway weaving sections, *Transportation Research Record*, **1398**, Transportation Research Board, Washington, USA, 42–48.

Pipes, L. A. (1967) Car-following models and the fundamental diagram of road traffic, *Transportation Research*, **1**, 21–29.

Prigogine, I. & Herman, R. (1971) *Kinetic Theory of Vehicular Traffic*, New York: American Elsevier.

Rees, T., White, J. & Quick, J. (2000) *Monitoring of the M4 bus lane*, Unpublished Report, PR/T/125/2000, Transport Research Laboratory, Crowthorne, UK.

Richards, P. I. (1956) Shockwaves on the highway, *Operations Research*, **4**(1), 42–51.

Ritchie, S. G. & Cheu, R. L. (1993) Simulating of freeway incident detection using artificial neural networks, *Transportation Research Part C*, **1C**(3), Pergamon Press, Oxford, 203–217.

Rose, G. & Dia, H. (1995) Freeway automatic incident detection using artificial neural networks, *Proceedings of the International Conference on Application of New Technology to Transport Systems*, **1**, May 1995, Melbourne, Australia, 123–140.

Seddon, P. A. (1972) Program for simulating dispersion of platoons in road traffic, *Simulation*, **18**, 81–90.

Stanton, P. L. (1992) The operation of grade separated junctions, *Proceedings of Seminar H, PTRC European Transport, Highways and Planning 20th Summer Annual Meeting*, London, 233–245.

Still, P., Abou-Rahme, N. & Rees, T. (1997) *The causes of congestion – final report for HA Project N510*, Confidential Report PR/TT/151/97, Transport Research Laboratory, Crowthorne, UK.

Taylor, N. (1998) *Current and emerging automatic incident detection algorithms*, Unpublished report PR/TT/163/98, Transport Research Laboratory, Crowthorne, UK.

Underwood, R. T. (1961) Speed, volume, and density relationships, quality and theory of traffic flow, *Yale Bureau of Highway Traffic*, 141–188.

Wang, M. H., Cassidy, M. J., Chan, P. & May, A. D. (1993) Evaluating the capacity of freeway weaving sections, *Journal of Transportation Engineering*, **119**(1), 48–54.

Wardrop, J. G. (1952) Some theoretical aspects of road traffic research, *Proceedings of the Institution of Civil Engineers*, Pt II,**1**, 325.

Wedlock, M., Peirce, J. & Wall, G. (2001) *A review of anti-swooping trials*, TRL Report **491**, Crowthorne, UK

White, J. (2001) *M25 business case – report on journey time research*, Unpublished Report 42553/DOC/2506, Mouchel Consultants Ltd, UK.

Whitham, G. B. (1999) Linear and non-linear waves, *Pure and Applied Mathematics Series*, Wiley, New York, 68–80.

Zarean, M. & Nemeth, Z. A. (1988) WEAVSIM: A microscopic simulation model of freeway weaving sections, *Transportation Research Record*, **1194**, Transportation Research Board, Washington, USA, 48–54.

BIBLIOGRAPHY

Holmes, B. (1994), When shockwaves hit traffic, *New Scientist*, **142**(1931), June 25, 1994 IPC Magazines, London, 36–40.

Lee, P. M. (1997) *Bayesian Statistics – An Introduction*, London: Arnold Books.

Leutzbach, W., & Wiedemann, R. (1986) Development and applications of traffic simulation models at the Karlsruhe Institut fur Verkehrwesen, *Traffic Engineering and Control*, **27**, 270–278.

Sonka, M., Hlavac, V. & Boyle, R. (1993) *Image processing, analysis and machine vision*, London: Chapman and Hall Computing.

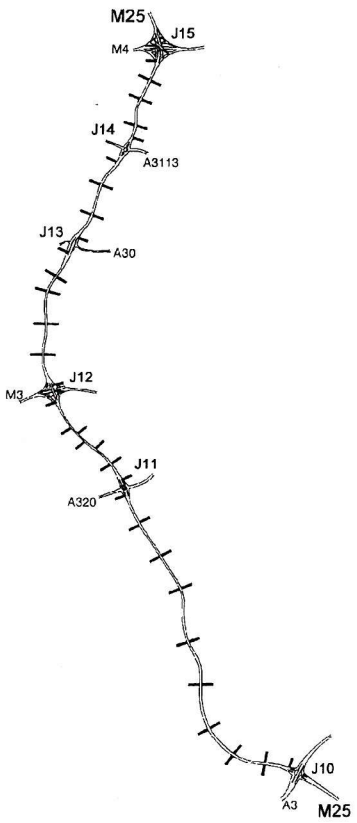
Persaud, B. N., Hall, F. L. & Hall L. M. (1990), Congestion identification aspects of the McMaster incident detection algorithm, *Transportation Research Record* **1287**, pp. 167–175.

Zhou, M. & Hall, F. L. (1999) Investigation of the speed-flow relationship under congested conditions on freeways, *Transportation Research Record* **1678**, pp. 64–72.

APPENDIX A
CONTROLLED MOTORWAYS

SCHEME OVERVIEW

The purpose of this appendix is to present further information about the controlled motorway scheme and associated monitoring activities. The author was project manager and lead researcher on a number of research contracts for TRL Limited during the study period. Material in this section is adapted from Abou-Rahme, Dixon and Harbord (2002).



The M25 Controlled Motorway Section extends from Junction 10 (A3) to Junction 15 (M4) along the south-western quadrant in both directions. The system consists of enforceable Variable Speed Limit displays mounted on gantries at intervals throughout the section (see Figure A1). Traffic information collected by means of inductive loops is processed by various algorithms in order to determine the most appropriate speed limit to be displayed.

Mandatory motorway signals were installed at the beginning of 1995, but were operated with advisory speed limits (no red rings) until the scheme was launched in August 1995. A six-month programme of "before" monitoring began in March 1995. A basic traffic data-collection system was used to collect the traffic data needed for subsequent system performance-analysis and to develop initial control strategies. The enforcement system installation was completed in May 1995, but during the period between May and August 1995 the camera enforcement 'warning' signs were displayed without automatic enforcement.

On 31 August 1995 the system was switched on between Junctions 11 and 15 under a fixed-time plan. On 12 September 1995 it moved to automatic flow-based operation, and on 8 November 1995 the system was enabled between Junctions 10 and 11.

Between September and December 1995, the system parameters were adjusted to provide stable operation, and from January to April 1996 the parameters were further adjusted to evaluate alternative control strategies and optimise traffic control. In February 1997 the system switched to a speed-flow algorithm, and for the remainder of the 12-month pilot the system operated with optimised parameters to allow an assessment of its effectiveness.

The system operation was monitored throughout the first year of the pilot and assessed to determine if there were any areas where improvements could be made. Monitoring activity consisted of evaluation and processing of traffic and signal data to determine performance and trends. Graphical techniques have been used in order to identify shockwaves and secondary accidents. Over the lifetime of the project, it has been possible to evaluate the effects of the variable speed limits on the motorway and the potential benefits of the system. Areas of interest have included accident reduction, the level of noise and air-pollution, journey times, and driver attitudes to the system. In addition, monitoring activities have assisted in the fine-tuning of thresholds and identification of potential problems and inconsistencies. Benefits relating to changes in traffic behaviour following compliance with the system have also been identified.

The success of the first year operation of the pilot meant that it continued in operation and has been further developed. The second year of operation has confirmed the high level of compliance with speed limits, indicating that their credibility has been maintained, and introduced further improvements to the system operation.

In October 1997 a further feature of MIDAS was introduced on the M25, to protect drivers at the back of queues. The MIDAS outstations use the HIOCC

incident detection algorithm to detect a queue or slow-moving traffic. The MIDAS system combines HIOCC and speed control alerts to determine which signals should be set in order to best manage the traffic and protect the queue.

Prior to its implementation an assessment of potential safety issues arising from the introduction of a 40mph setting to traffic previously controlled by the national speed limit was carried out. It was recommended that in such cases a 50mph speed limit should be displayed for at least 5 seconds before the 40mph was displayed.

TRAFFIC DEMAND AND PERFORMANCE

Annual Average Daily Traffic (AADT) one-way flows for each link and carriageway have been calculated from MIDAS data. Figure A2 and A3 show the AADT flows along two sample links (Junctions 10 to 11 clockwise and Junctions 14 to 13 anticlockwise) for Aug 1995 to Dec 1998.

There are consistent seasonal trends for each site, with flows dropping during the winter and reaching a peak during July. On the A-Carriageway high summer flows are sustained during August, whilst on the B-Carriageway flows drop after July. There is a yearly growth of between 3% and 4%, with greatest increase on the section that had least traffic, namely Junctions 10–11.

Bidirectional AADT flows for M25 J10–15 are as follows:

	1996	1997	1998
J10–11	149,000	155,000	162,000
J11–12	163,000	169,000	176,000
J12–13	169,000	174,000	179,000
J13–14	183,000	187,000	191,000
J14–15	172,000	177,000	182,000

Table A1 – AADT Flows (Two-Way)

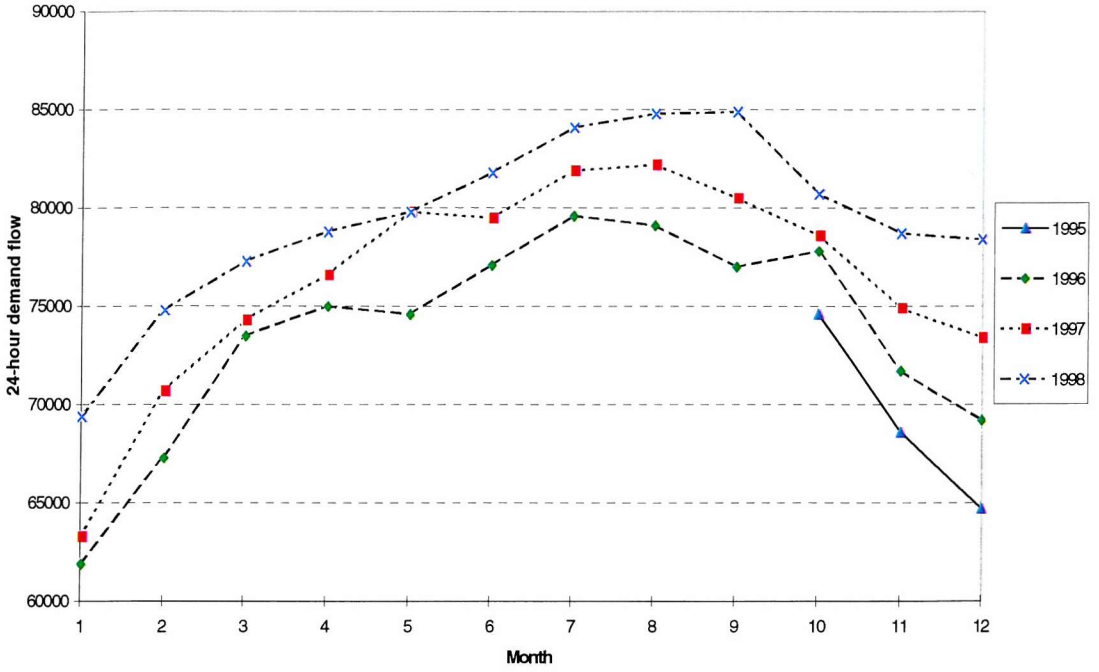


Figure A1 – AADT Flows on A-Carriageway (Junction 10–11)

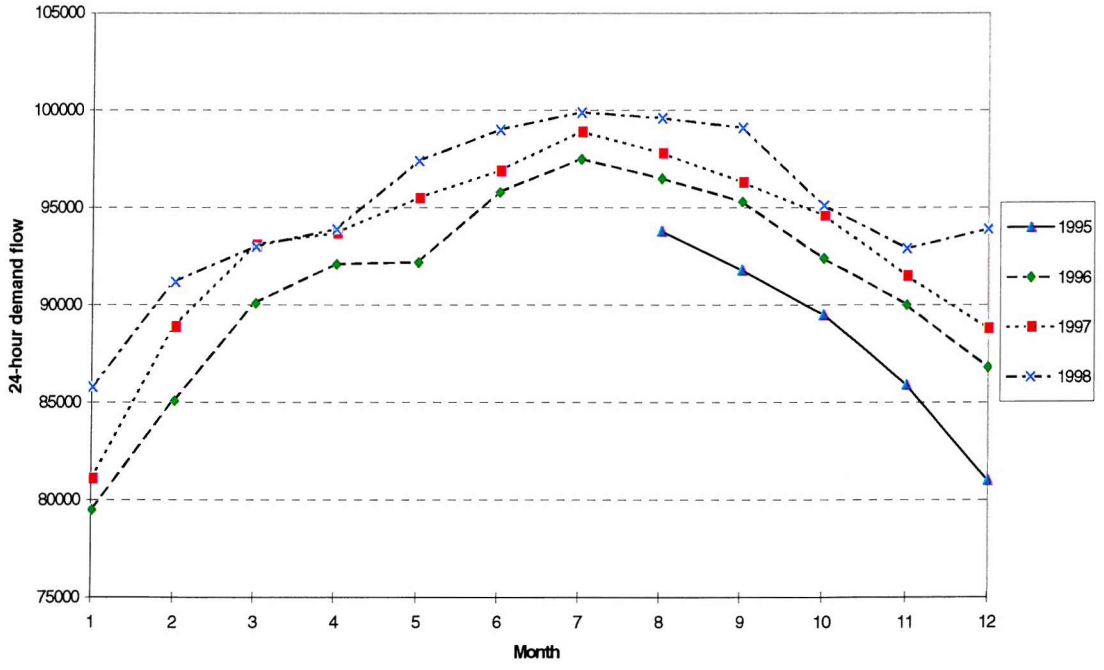


Figure A2 – AADT Flows on B-Carriageway (Junction 14–13)

In addition to the AADT flows, Annual Average Weekday Traffic (AAWT) 24-hour flows have also been analysed. AAWT flows show the same patterns as the AADT flows, but are greater in magnitude for each carriageway reflecting the fact that flows are greater during weekdays than on weekends. The AAWT 24-hour flows (two-way) are:

	1996	1997	1998
J10–11	156,000	163,000	171,000
J11–12	174,000	180,000	187,000
J12–13	180,000	185,000	190,000
J13–14	197,000	200,000	203,000
J14–15	184,000	188,000	194,000

Table A2 – AAWT 24-Hour Flows (Two-Way)

COMPLIANCE WITH SPEED LIMITS

Compliance with posted speed limits is an important test of system performance, since it is from this compliance that most of the other benefits of the system are expected to arise. Direct control over speeds is exercised only during the shoulders of the peak periods when the vehicles would be capable of travelling at speeds in excess of the speed limits displayed.

The introduction of the variable speed limits in August 1995 has had a significant effect on speeds at which the drivers travel, (even when speed limits are not displayed). The number of speeding drivers has dropped, with greater improvement observed for very high speeds. The percentage of drivers exceeding 70mph has dropped by 17% and the percentage exceeding 90mph has dropped by 46%. By comparison, the introduction of camera warning signs in May 1995 had little effect on driver behaviour.

Comparison of speed-profiles during similar periods in consecutive years suggests that the 40mph speed limits have reduced the extent to which drivers can overshoot the optimum speed for local traffic conditions following a shockwave. This can be interpreted as a reduction in the amount of braking

and over-accelerating, and the perception of drivers passing through the section will be that of a smoother ride. The option to display a more appropriate speed limit also contributes to the overall credibility of the system. This evidence points to a continued compliance with the mandatory speed limits.

Examining the characteristics of the shockwaves over three years, there is little evidence to suggest any change either in the minimum speed experienced within them, or in the frequency of their occurrence. This stability should be considered in the light of the 3% increase in demand per annum for the controlled motorway section.

HEADWAY ANALYSIS

Another objective of the controlled motorway scheme is to improve headway distribution within each of the lanes, reducing both the risks associated with sudden braking, as well as the stress of driving. It was found that the number of very short headways had decreased within all the pre-defined speed bands. As speeds increase, the number of short headways also increases, implying that the gap some drivers leave between vehicles is based on distance, rather than time.

At speeds between 50 and 60 mph, 4% of vehicles are less than 0.7 seconds behind the vehicle in front, equivalent to a gap of about 13 metres. Overall there are fewer short headways and the distribution is more uniform. Average headway has also been reduced implying increased efficiency in road operation. Significant improvements in the speed band of 40–50mph are attributable to the controlled motorway system.

SYSTEM PERFORMANCE

The introduction of a 40mph speed limit carries a measure of risk when displayed to traffic not previously controlled by a speed limit below 70mph. An investigation found that when a 50mph speed limit is displayed first the risk of collision is greatly reduced. The current system displays a 50mph speed limit before switching to a 40mph speed limit after 5 seconds.

The automatic incident-detection system was implemented during October 1997. Signal sequences provide adequate protection for both shockwaves, and traffic queuing due to an incident downstream.

TRAFFIC BEHAVIOUR

Average demand flows (both AADT and AAWT) have been increasing by approximately 3% per year along the controlled section. Examination of traffic flows for the morning peak period during consecutive Mondays over two years provides some evidence of peak shifting. At site 4902A where flow-breakdown often occurs first along the Controlled Motorway section there was an increase in flows just after 06:00. Examination of equivalent flows for site 4811A near Junction 11 indicated that the start of the morning peak period has moved forward by a few minutes every year.

The passage of a slow vehicle outside the peak period (often accompanied by a police escort) can act as a moving seed-point. The behaviour of drivers following at slow speed leads to the propagation of a number of shockwaves in the wake of the slow vehicle. The vehicle-hour delay arising from the passage of a slow vehicle can be of the same magnitude as the delays arising during a peak period.

The 'rubbernecking' phenomenon is defined as when a speed-drop or flow-breakdown occurs on one carriageway as a result of an incident or distraction on the opposite carriageway. Whilst rubbernecking does not occur very frequently it can sometimes double the amount of congestion arising from a single incident. The example of 23 January 1998 shows that flows need to be sufficiently high before a disturbance due to a distraction on the opposite carriageway will actually lead to flow-breakdown.

The recurrence of a seed-point downstream of Junction 12 on the anticlockwise carriageway during the evening peak period has been identified. Another site where flow-breakdown is becoming increasingly noticeable is at the Junction 14 merge on the clockwise carriageway between 16:00 and 21:00 on a Sunday afternoon. The extent of the congestion during this period is as severe as that experienced during the weekday morning peak period at

the same location. Along the clockwise carriageway, the merge at Junction 11 and the midsection at Junction 13–14 remain the two most common sites for flow-breakdown. At the beginning of the morning peak period the first shockwaves occur at either one of these sites.

SPECIFIC STUDIES

Monitoring work also comprised a number of short studies into specific areas.

A driver opinion survey was conducted to determine the impact of the scheme on drivers who use it. The survey yielded favourable results with more than two-thirds of respondents saying they would like to see the system extended in some way.

Benefits from the reduction in noise levels and air pollution are attributable to the impact of speed-limit enforcement. Studies indicated that most benefits are likely to occur at the shoulders of the peak period or during periods when the signs are set and flow-breakdown has not yet occurred.

The specific effect that signal trials might have had on injury accidents was examined by comparing accident statistics from five years before and two years after the commencement of the pilot. Initial regression analysis showed a net reduction in injury accidents on the trial section of around 12% during the period of signal operation. The large confidence intervals introduced by filtering the effects of lighting, roadworks and number of lanes made it difficult to ascertain the *specific* effect of the signal trials on injury accidents.

SUMMARY FOR CONTROLLED MOTORWAYS

The system provides a beneficial environment for drivers using the M25, absorbing the effects of a 3% increase in demand per annum. An automatic incident-detection system has been in place since October 1997, detecting, tracking and protecting the back of a queue during both peak and off-peak periods. The EMS serves to provide additional information to drivers regarding the reason for a particular signal display.

APPENDIX B
DATA COLLECTION METHODS

PRACTICAL CONSIDERATIONS

Description of Alternative Approaches

During the period of this research the author has been responsible for the design and implementation of a number of motorway evaluation projects, most notably the M25 Controlled Motorway Section. At the time of study, the primary source of available data was from the MIDAS induction loop system. A number of alternative technologies were available so these had to be considered in terms of what this study was seeking to achieve. At the time of completing this thesis (2003), the author is aware of a number of new studies commissioned by the Highways Agency to look at high-tech alternative data collection methods (including fibre-optic and infrared detectors) but these do not impact upon the study methods used here.

The acquisition of good quality information on speed, flow, and headway was of particular interest. MIDAS provided only averaged information and the data had a spatial resolution of around 500m. A number of possible alternative approaches to data collection were considered as follows:

Video Survey (Mounted Cameras): A number of cameras would need to be mounted on the central reservation in order to obtain an accurate picture over at least a 100m length, without being obscured by heavy goods vehicles. However, during playback a number of problems are introduced. Firstly, with regard to resolution, the number of lines on a monitor is around 600. Assuming the view to be side-on, it would not be possible to determine the headway to a satisfactory degree of accuracy. The best possible scenario would be an overhead view (like CCATS) making full use of each line. In order to capture a 100m length, the resolution would be 0.16m per line, but the

errors from parallax due to the monitor glass in trying to measure the distances would reduce accuracy to 0.5m or worse. A refresh rate of 25 times per second is good resolution, but is obscured by the loss in spatial resolution.

Fast Reel Film: Ideally, this approach would use an automatic camera and cartridge that could take 250 exposures at a rate of 5 frames per second. The graphic resolution of the camera would be correct to around 0.1m at the distance being considered here. Two significant problems are (a) the discrete and effective mounting of the camera, and (b) determining when to film in order to catch a shockwave in its early stages. In addition to this, whilst the camera resolution would be accurate to 0.1m the measurement techniques (using an overlay grid to determine distances) would reduce the accuracy to below 0.5m after errors have been accounted for.

Portable Loops: This would provide an improved resolution on the section of motorway being considered, although the quality of information relayed by such loops might be limited at lower prices. There were also problems with obtaining permission to install, and the expense was considered prohibitive. An alternative 'detector' approach would involve using sodium lamps in conjunction with the existing cats-eyes, but this is even more expensive for the desired resolutions.

Mini Aircraft: The advantage of such devices is the fact that once airborne they are virtually undetectable to the human eye and so there is no distraction. There is a difficulty with the amount of time such a device would remain in view of the motorway, not to mention the cost of such an operation. Alternatives include mounting a camera on a hot-air balloon, but despite a number of contacts it has not been possible to secure permission for filming.

Instrumented vehicles: This may require a number of vehicles to pass through a section of turbulent flow in order to detect the perturbations and to understand inter-vehicular behaviour. Instrumented vehicles are available through Southampton University, so would have been a feasible form of data collection. A number of drive cycle surveys were commissioned by the author

for the M25 J15–16 Controlled Motorway Business Case project, and the data was used to help with calibration of speeds (and corresponding journey times) obtained through MIDAS. On reflection, these drive cycles offer useful information about the propagation of shockwaves but are not particularly helpful for seed-point research.

GPS Receptors: A variation of the above idea, this approach would involve connecting GPS receptors to a number of vehicles in order to collect positional information as the vehicle drove through the section of interest. However, even with differential GPS and the software set to maximum update, the resolution is still outside the acceptable limits.

Video Survey (Hand-held in vehicle): This is an alternative to the external mounted cameras. It might be possible to take a camera in the drive-through vehicle, filming the brake-light pattern ahead. The camera operator would also need to be able to identify all the marker posts and record a commentary on the driving conditions. This approach only provides a limited view, and it would be difficult to quantify the headway involved, even if the speedometer was included in the view for speed measurements.

Equip vehicles with beacon: Cars would be able to store speed data although this is likely to be averaged before it is available for processing. Seven cars (excluding the two lab vehicles) are available under various projects. Problems with this approach include the need for an accurate synchronisation of the vehicles as they pass the beacon.

Driver Behaviour: As a parallel study, it would be necessary to test different types of driver behaviour. One approach would be to bring in a variety of drivers who use the section on a regular basis and conduct a 'talk through' experiment. Two particular aspects of behaviour will be tested, namely to talk through the actions as they drive and to construct a profile of the behavioural type. This approach is a variation on the use of instrumented vehicles.

MIDAS Loop data: Current resolution is approximately every 500m but the data is also averaged over one-minute intervals. However, it is possible to obtain similar information without averaging, so that an actual speed, flow and occupancy data set can be collected. This type of data is referred to as IVD.

Conclusion on Alternative Approaches

A reasonable assumption of the deceleration rate of a vehicle without braking is around 3kph/s (i.e. 0.8 m/s^2). For a subtle drop in speed, say from 50mph (22.2m/s) to 48mph (21.1m/s), the time to achieve this would be around 1.4 seconds over a distance of around 30m. If we consider a frame by frame video analysis working at say one frame every second (being twice the rate of driver reactions) and consider the possibility of a 0.4m random error being introduced, the experiment is no longer feasible. This is because the total error between frames (0.8m over a 1s interval for an approximate speed drop of 1m/s) is as large as the acceleration being investigated! Reducing the time interval would increase the impact of the error.

Of the approaches discussed here, some appear to be of low feasibility and others appear to be of little use. The motorised camera appears to be the most satisfactory approach for obtaining an accurate snapshot, but is not without difficulties. With regard to the drive-through approach, such data could be better applied to understanding the propagation mechanism (since the vehicles could experience a number of shockwaves during a drive through the Controlled Motorway section) instead of trying to 'catch' the starting point.

This review of data collection methods informed the decision to begin working with the existing MIDAS loop data, and took the focus of the study away from the perturbations which cause flow-breakdown, and onto the conditions prevalent at the time of flow-breakdown.

APPENDIX C

GIPPS CAR FOLLOWING ALGORITHM

ASSUMPTIONS

The purpose of this appendix is to present the Gipps Car Following Algorithm for easy reference. The notation of the original paper is used:

- a_n maximum acceleration chosen by the driver of vehicle n,
- b_n most severe braking applied by the driver of vehicle n (and $b_n<0$),
- s_n effective size of vehicle n, comprising physical length and a margin into which the following-vehicle does not encroach even at rest,
- V_n speed at which the driver of vehicle n wishes to travel,
- $x_n(t)$ location of the front of vehicle n at time t,
- $v_n(t)$ speed of vehicle n at time t,
- τ apparent reaction time, a constant for all vehicles.

Gipps made three key assumptions in formulating his algorithm:

(A1) Actual vehicle speed will not exceed the desired speed at any time

(A2) Free acceleration should increase with speed (as engine torque increases) then decrease to zero when approaching its desired speed

(A3) A driver will at all times maintain sufficient headway with the vehicle in front. If the lead-vehicle stops as quickly as possible and it takes the following-driver time τ to react, the following-vehicle can still stop without collision and without exceeding the maximum desired braking rate.

Assumptions (A1) and (A2), applied to vehicle n, are summarised by the single inequality:

$$v_n(t+\tau) \le v_n(t) + 2.5a_n\tau\left(1-\frac{v_n(t)}{V_n}\right)\sqrt{\left(0.025+\frac{v_n(t)}{V_n}\right)}. \qquad (1)$$

This equation is used by Gipps to describe **free-flowing traffic** and was obtained by examining the speed of an instrumented vehicle driving through moderate traffic and deriving an appropriate relationship. In this model, a

vehicle in free-flowing traffic is influenced by three factors all relating to itself, namely desired speed, previous speed and maximum acceleration. Assumption (A3), applied to vehicle *n*, is described by the inequality:

$$x_n^* \le x_{n-1}^* - s_{n-1} \qquad (2)$$

where x_{n-1}^* is the position at which vehicle *n-1* will come to rest if it starts braking as hard as is desirable at time *t*, and x_n^* is the position at which vehicle *n* will come to rest if it starts braking as hard as is desirable at time $(t+\tau)$, a reaction time τ later.

APPLICATION

Using a formula of the type $v^2 - u^2 = 2as$ to relate acceleration, velocity and distance travelled, equation (2) becomes

$$x_{n-1}(t) - \frac{v_{n-1}(t)^2}{2b_{n-1}} - s_{n-1} \ge x_n(t) + [v_n(t) + v_n(t+\tau)]\frac{\tau}{2} + v_n(t+\tau)\theta - \frac{v_n(t+\tau)^2}{2b_n} \qquad (3)$$

where θ is an extra reaction time term. Without this extra term θ, a vehicle approaching a stationary vehicle or a stop line would continue to travel at its desired speed until the minimum headway was reached, at which point the driver would brake at its maximum desired braking rate. The effect of adding this term is to cause vehicles to commence braking earlier, and to reduce braking as they approach the point at which they must stop, and thus creep slowly up to this point.

The solution of the above inequality, for $v_n(t+\tau)$ is given by the positive root of Equation (3),

$$v_n(t+\tau) \le b_n\left(\theta + \frac{\tau}{2}\right) + \sqrt{b_n^2\left(\theta + \frac{\tau}{2}\right)^2 + \frac{b_n}{2}\left(v_n(t) + v_{n-1}^2(t)/\hat{b} - 2(x_{n-1}(t) - s_{n-1} - x_n(t))\right)} \qquad (4)$$

For car-following behaviour, the speed of the vehicle is now influenced by characteristics of the vehicle ahead as well as its own properties. The speed, maximum braking, effective vehicle length and displacement of the vehicle ahead determine the new speed of the following vehicle.

Equations (1) and (4) represent two constraints on the speed of vehicle n at time $t+\tau$. Assuming a driver travels as fast as safety and the limitation of the vehicle permits, the new speed is given by the minimum of the right hand sides of Equations (1) and (4). The displacement of the vehicle is calculated as follows,

$$x_n(t+\tau) = x_n(t) + \frac{\tau}{2}(v_n(t) + v_n(t+\tau)) \qquad (5)$$

The Inter-Vehicular distance (IV) is defined as

$$IV(t) = x_{n-1}(t) - x_n(t)$$

If this distance is smaller than the actual vehicle length of the vehicle ahead then a collision will have taken place, and when testing the parameters this scenario should be avoided.

According to Gipps, the behaviour of the vehicles is determined, to some extent, by the relative magnitudes of θ and τ. When θ is equal to $\tau/2$ then a vehicle travelling at a safe speed and distance will be able to maintain a state of safety indefinitely providing the willingness of the driver ahead to brake hard has not been underestimated. In real traffic, this is not always the case and therefore when the willingness of the driver ahead to brake has been underestimated, θ may have to be increased to avoid vehicles colliding. With θ equal to $\tau/2$, the following regimes exist. Congested flow exists when Equation (4) is the limiting condition. Then all vehicles flow as fast as the volume allows. When Equation (1) is the limiting condition for a substantial proportion of the vehicles, free flow is obtained.

A smooth transition is possible between the two conditions. This does not occur when the lead vehicle brakes harder than the following vehicle has anticipated, or leaves the lane, or when a new vehicle moves into the gap between two vehicles.

The model, with θ equal to $\tau/2$, has been validated using values taken through statistical sampling. Gipps introduces a term $\hat{b}$ with little explanation. It appears to limit the maximum value of the estimated braking parameter to –

$3m/s^2$. An interesting application of this value is the effect it has on disturbances to the traffic flow.

When $\hat{b}$ is less than b_{n-1} the term containing the speed of the vehicle ahead contributes *more* to the new speed of the vehicle, and the disturbances to the following vehicles are damped. This means that the driver has *overcompensated* for the velocity of the vehicle ahead. However, when the braking of the driver ahead is underestimated, these disturbances could be amplified. This also means that a smooth transition between the two possible values for $v_n(t+\tau)$ does not occur.

In summary, Gipps postulates that three factors control the behaviour of the traffic,

- The distribution of desired speeds
- The average reaction time for drivers
- The ratio of mean braking to the driver estimate of mean braking

APPENDIX D
CALIBRATING MIDAS

PURPOSE

The purpose of this appendix is to illustrate the nature and scale of possible inaccuracies in the detector readings for MIDAS. This is done in two parts, (i) a comparison with an ANPR survey for M25 and (ii) a comparison with the TITAN Database for M6 Data.

COMPARISON WITH ANPR

Abou-Rahme, White and Stewart (2002) present the findings of research conducted for Mouchel Consulting whilst deriving a controlled motorway business case. Further details are contained in White (2001).

The time-mean speed is an arithmetic average of spot speeds measured at a single detector point, an example of which is MIDAS. The space-mean speed is measured over a fixed distance using time difference, an example of which could be derived through ANPR Camera timestamps.

In free-flow conditions, the time-mean speed is a good representation of the average speed of the vehicles that have passed over the loop in one minute. However, in congested conditions, where speeds may vary by large amounts over each minute, the time-mean speed overestimates the actual speed of vehicles in that minute.

Data was initially collected for one week in December 2000, but then a continuous set of data was collected over the period 30 January 2001 to 02 April 2001, from 8 loops on the section. Comparisons were made between journey times calculated from MIDAS and ANPR, leading to an improved approximation of true journey times using a proxy for space-mean speed.

Figure D1 is reproduced from that work, and reflects the fact that there are two solutions to the space-mean speed equation for each value of time-mean

speed examined. According to Wardrop (1952) an approximate relationship between time-mean speed and space-mean speed can be made as follows:

$$\overline{v}_t = \overline{v}_s + \frac{\sigma_s^2}{\overline{v}_s}$$

The solution of this equation for space-mean speeds is therefore:

$$\overline{v}_s = \frac{\overline{v}_t \pm \sqrt{\overline{v}_t^2 - 4\sigma_s^2}}{2}$$

Figure D2 demonstrates the journey times obtained using time-mean speeds and space-mean speeds to calculate journey times for 27 February 2001 on M25 J15–16 over a distance of 6km. Calculations are based on individual vehicle data. It can be seen that the deviation during congested conditions is about 4%.

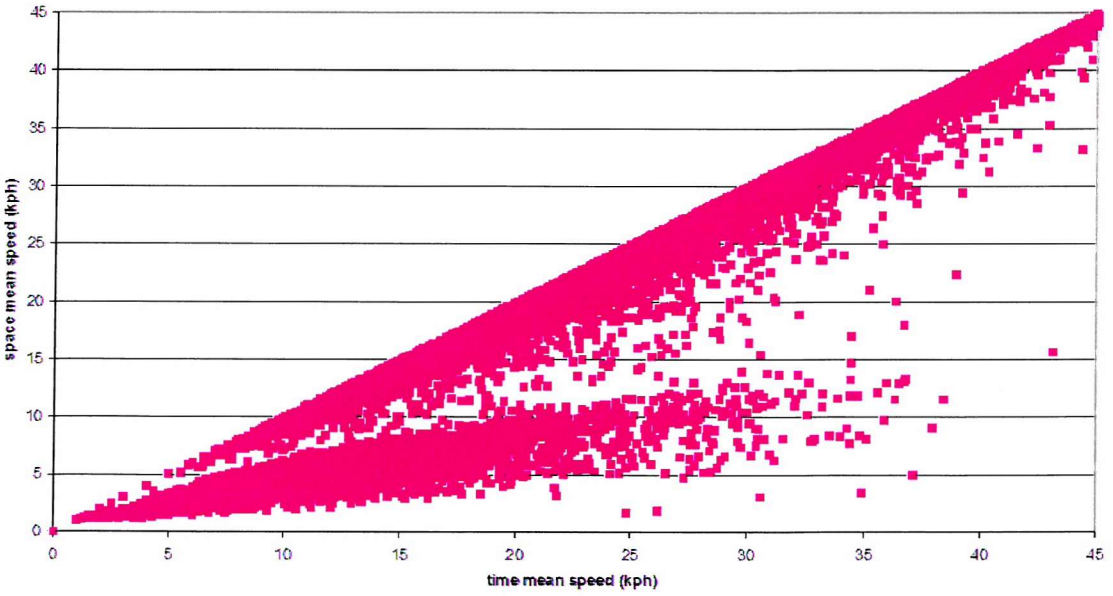


Figure D1 – Journey time comparison using time-mean and space-mean speeds

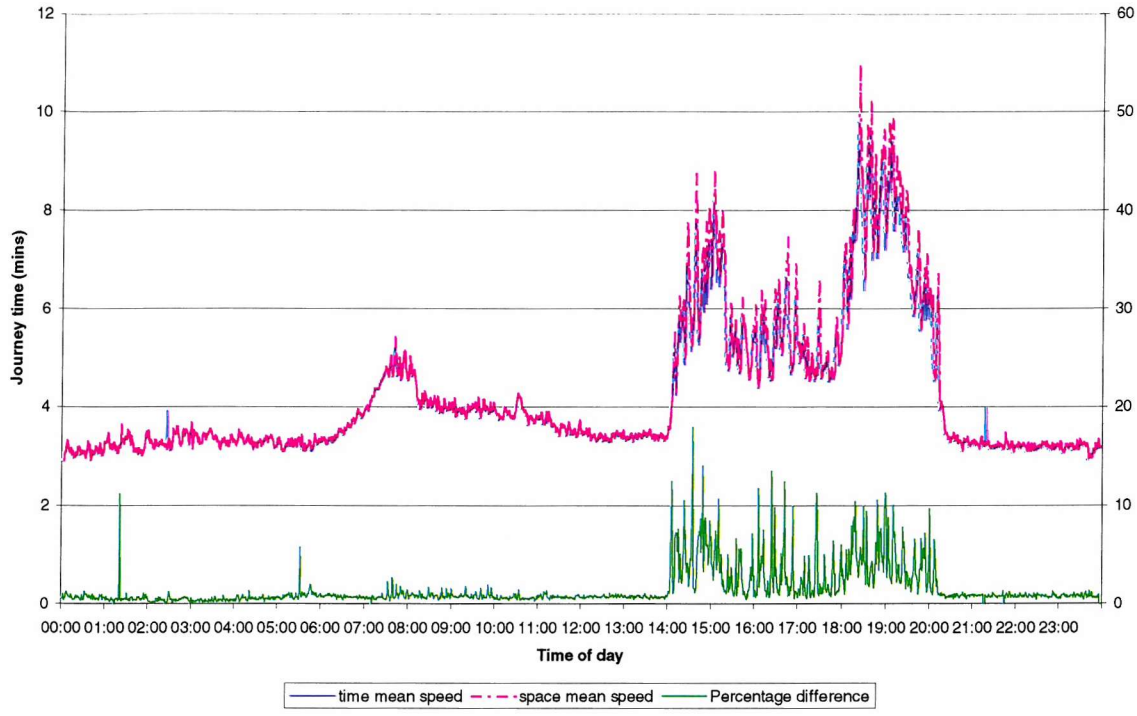


Figure D2 – Journey time comparison using time-mean and space-mean speeds

The journey-time study also looked at the relative values derived from ANPR Cameras and from MIDAS. The resulting deviations may have been attributable to approximations in the journey-time algorithm for MIDAS, the possibility of overestimating speeds during congested conditions, or indeed the accuracy of marker post positions on the carriageway. Figure D3 and Figure D4 illustrate the findings.

Figure D3 – Comparison of journey times from ANPR and MIDAS (22 February 2001)

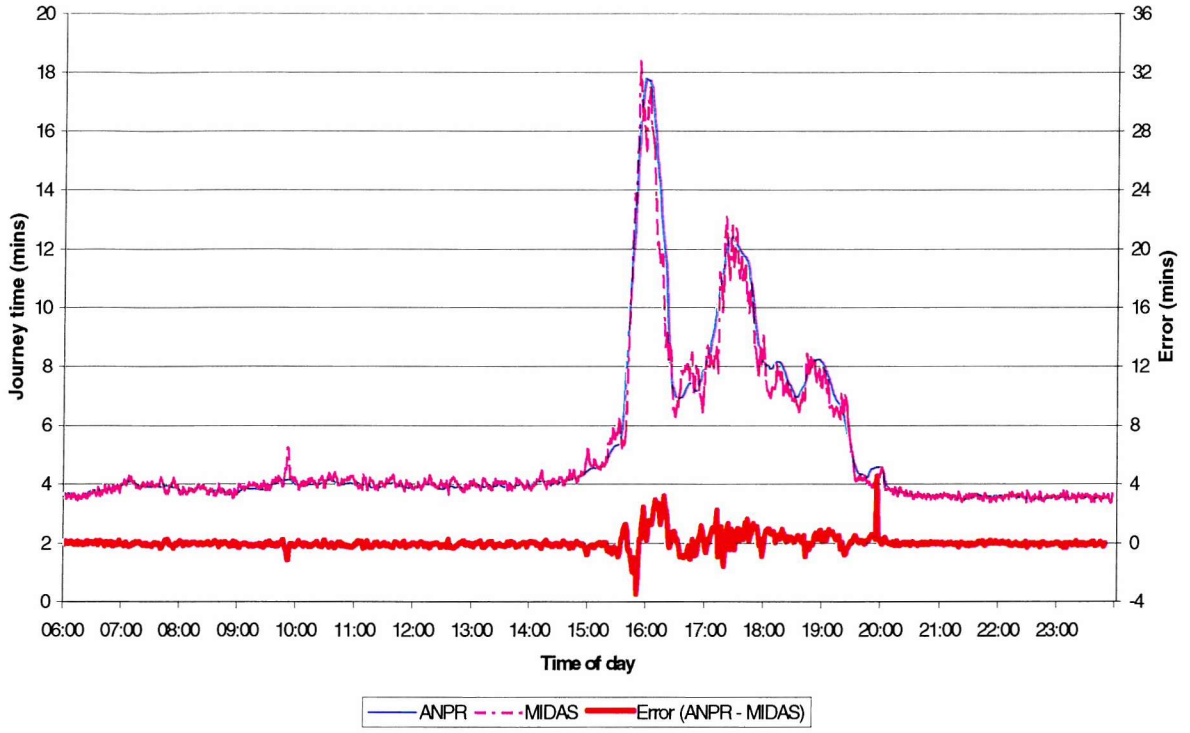
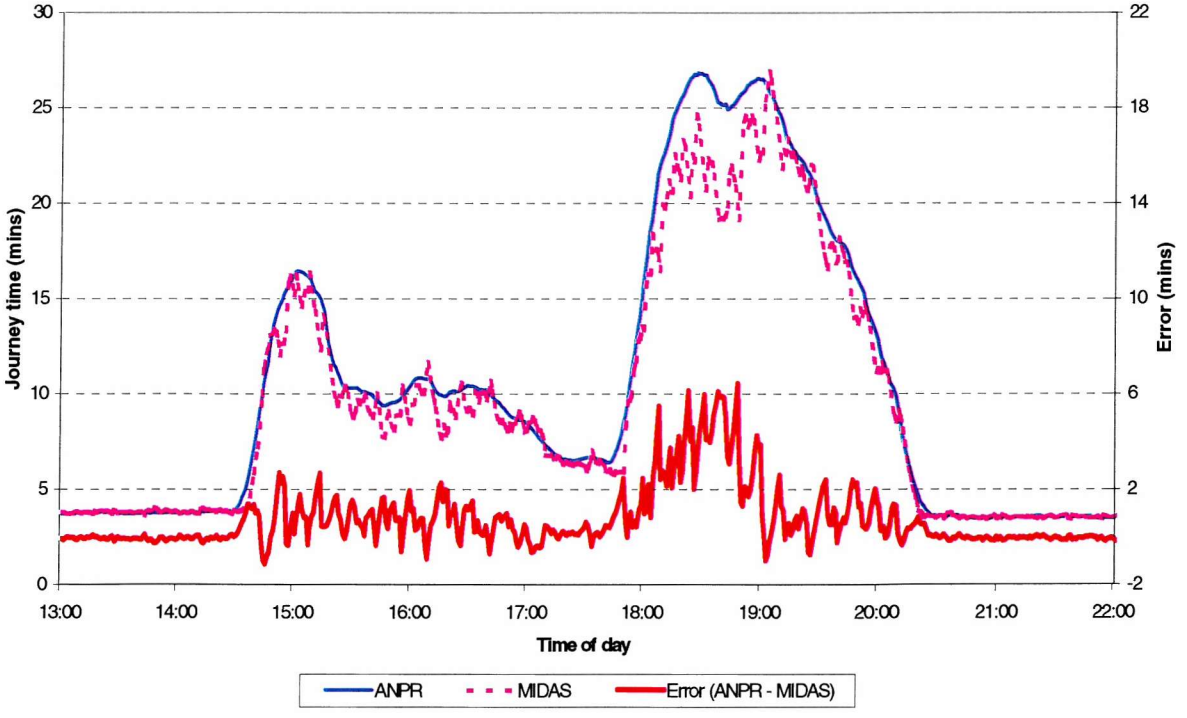


Figure D4 – Comparison of journey times from ANPR and MIDAS (27 February 2001)

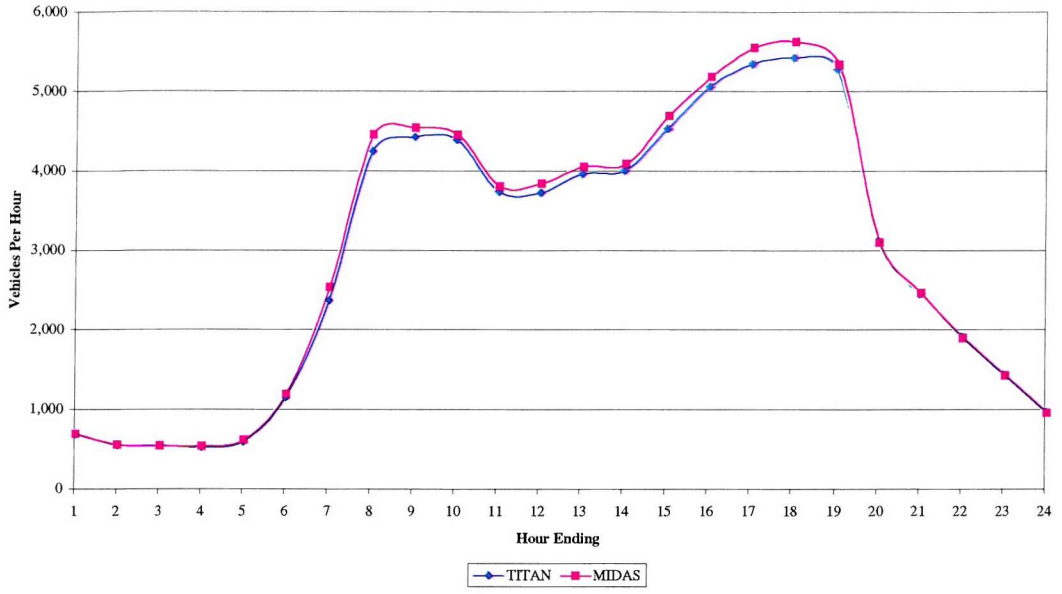


COMPARISON WITH TITAN

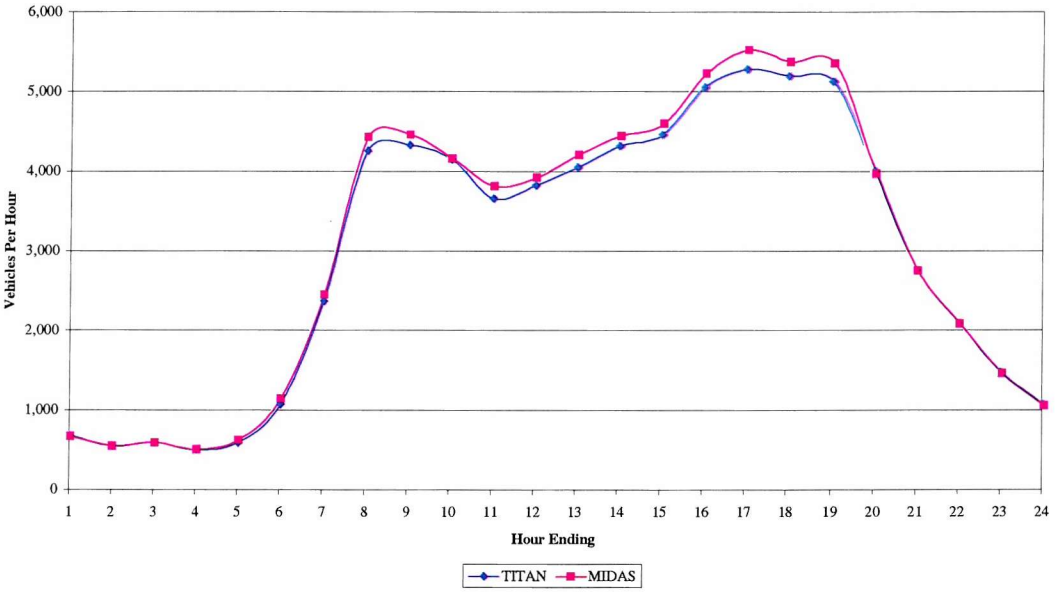
During the course of study, a comparison of MIDAS Calibration against TITAN on the M6 Northbound was obtained from the Highways Agency. This has been included below for reference along with graphs based on the table.

	Wednesday 20/01/99				Thursday 21/01/99				Friday 22/01/99			
			Difference				Difference				Difference	
Hour Ending	TITAN	MIDAS	Abs.	%	TITAN	MIDAS	Abs.	%	TITAN	MIDAS	Abs.	%
1	691	696	5	0.7	681	672	-9	-1.3	719	724	5	0.7
2	554	559	5	0.9	549	554	5	0.9	575	580	5	0.9
3	551	548	-3	-0.5	596	597	1	0.2	598	613	15	2.5
4	529	544	15	2.8	503	510	7	1.4	517	531	14	2.7
5	597	617	20	3.4	593	627	34	5.7	587	602	15	2.6
6	1,150	1,194	44	3.8	1,077	1,148	71	6.6	1,061	1,119	58	5.5
7	2,371	2,538	167	7.0	2,371	2,457	86	3.6	2,144	2,273	129	6.0
8	4,247	4,458	211	5.0	4,262	4,435	173	4.1	3,779	3,938	159	4.2
9	4,426	4,545	119	2.7	4,334	4,464	130	3.0	3,744	3,719	-25	-0.7
10	4,393	4,457	64	1.5	4,153	4,166	13	0.3	3,683	3,825	142	3.9
11	3,743	3,809	66	1.8	3,661	3,817	156	4.3	3,496	3,634	138	3.9
12	3,726	3,847	121	3.2	3,825	3,926	101	2.6	3,972	4,087	115	2.9
13	3,966	4,059	93	2.3	4,054	4,216	162	4.0	4,407	4,531	124	2.8
14	4,006	4,088	82	2.0	4,321	4,452	131	3.0	4,961	5,170	209	4.2
15	4,532	4,700	168	3.7	4,462	4,603	141	3.2	4,671	4,785	114	2.4
16	5,063	5,193	130	2.6	5,060	5,238	178	3.5	4,883	5,169	286	5.9
17	5,356	5,559	203	3.8	5,286	5,530	244	4.6	4,897	5,037	140	2.9
18	5,429	5,630	201	3.7	5,201	5,383	182	3.5	4,972	5,133	161	3.2
19	5,291	5,346	55	1.0	5,134	5,365	231	4.5	4,200	4,189	-11	-0.3
20	3,121	3,102	-19	-0.6	4,008	3,977	-31	-0.8	4,113	4,078	-35	-0.9
21	2,458	2,467	9	0.4	2,760	2,757	-3	-0.1	3,593	3,577	-16	-0.4
22	1,920	1,903	-17	-0.9	2,097	2,093	-4	-0.2	2,894	2,900	6	0.2
23	1,445	1,432	-13	-0.9	1,483	1,469	-14	-0.9	2,022	1,980	-42	-2.1
24	981	962	-19	-1.9	1,081	1,063	-18	-1.7	1,156	1,136	-20	-1.7
Total	70,546	72,253	1,707	2.4	71,552	73,519	1,967	2.7	71,644	73,330	1,686	2.4

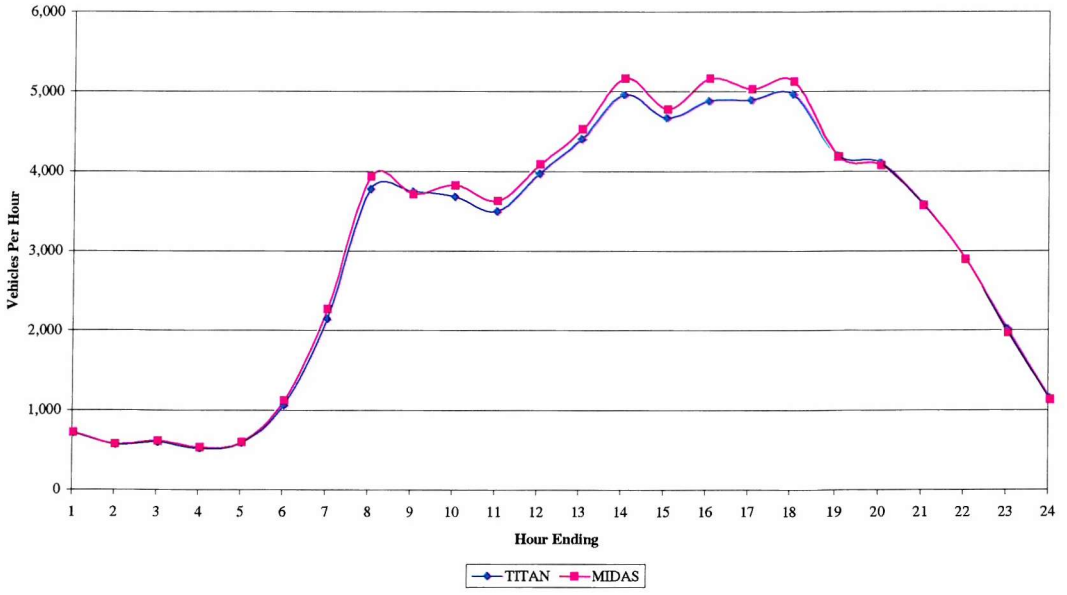
M6 J10-J10a NB - Wed 20/01/99



M6 J10-J10a NB - Thu 21/01/99



M6 J10-J10a NB - Fri 22/01/99



APPENDIX E

PROBABILITY ANALYSIS – ADDITIONAL PLOTS

This appendix presents supplementary plots in support of the results found in Chapter 6.

A summary of the statistical parameters used to create the bivariate normal distributions in Figure E1 and Figure E2 is given in Table E1 and Table E2 below.

Table E1 – Parameters for bivariate normal distribution, derived from June 2001 data

		μ_v	μ_k	$(\sigma_v)^2$	$(\sigma_k)^2$	cov (v,k)	$\rho_{v,k}$	sample
APS1	Rules 1-5 (f_1)	27.54	28.05	6.49	7.41	-0.36	-0.05	786
	Rules 1-2 (f_0)	29.47	26.68	15.91	6.32	-4.06	-0.40	4751

Table E2 – Parameters for bivariate normal distribution, derived from September 2001 data

		μ_v	μ_k	$(\sigma_v)^2$	$(\sigma_k)^2$	cov (v,k)	$\rho_{v,k}$	sample
APS1	Rules 1-5 (f_1)	27.57	27.72	6.38	6.44	-0.44	-0.07	3698
	Rules 1-2 (f_0)	28.74	26.62	14.38	6.39	-3.53	-0.37	25677

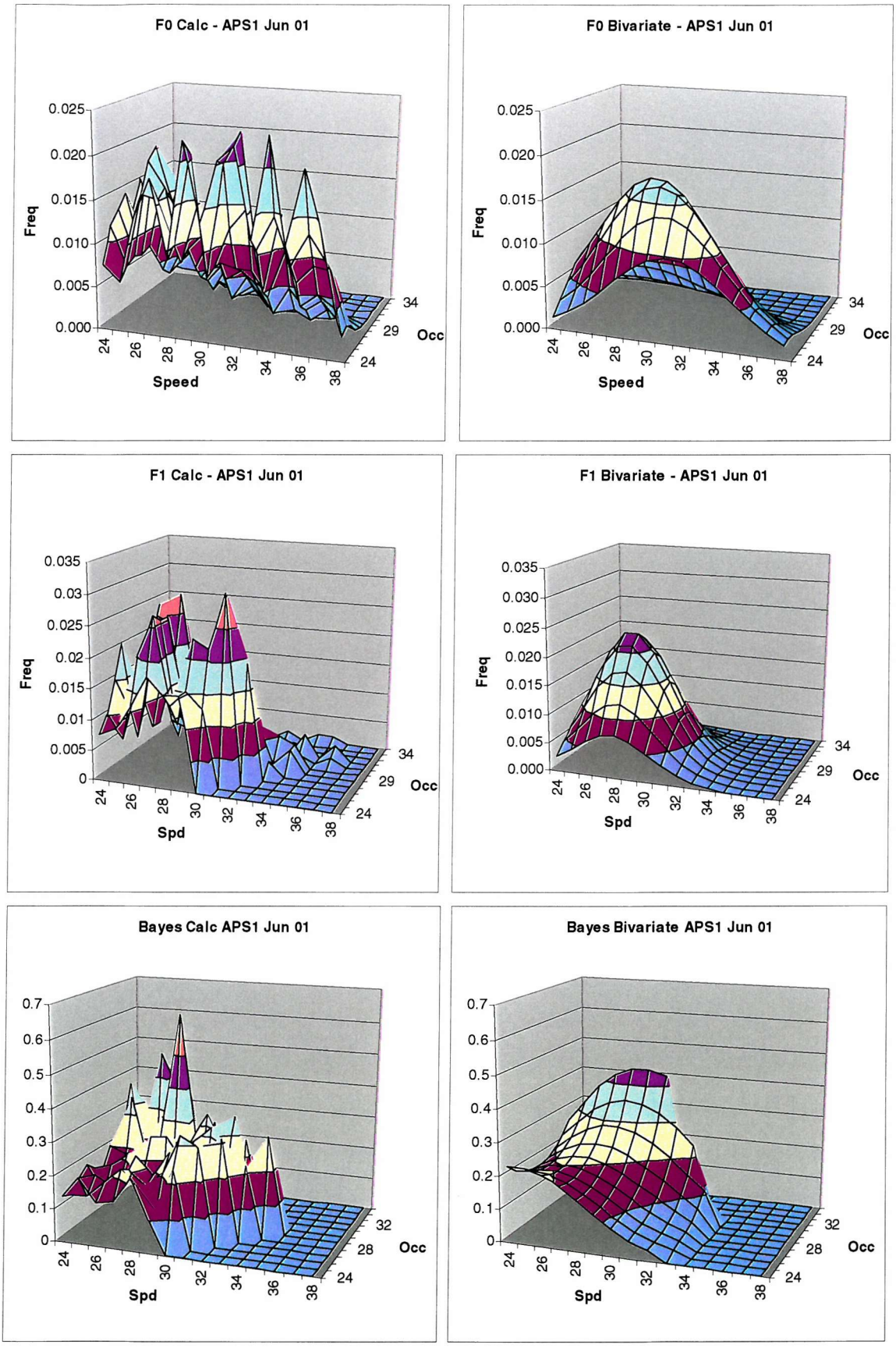


Figure E1 – Comparative plots for June 2001, showing distributions for f0, f1 and resulting posterior (using a uniform ratio for the prior function)

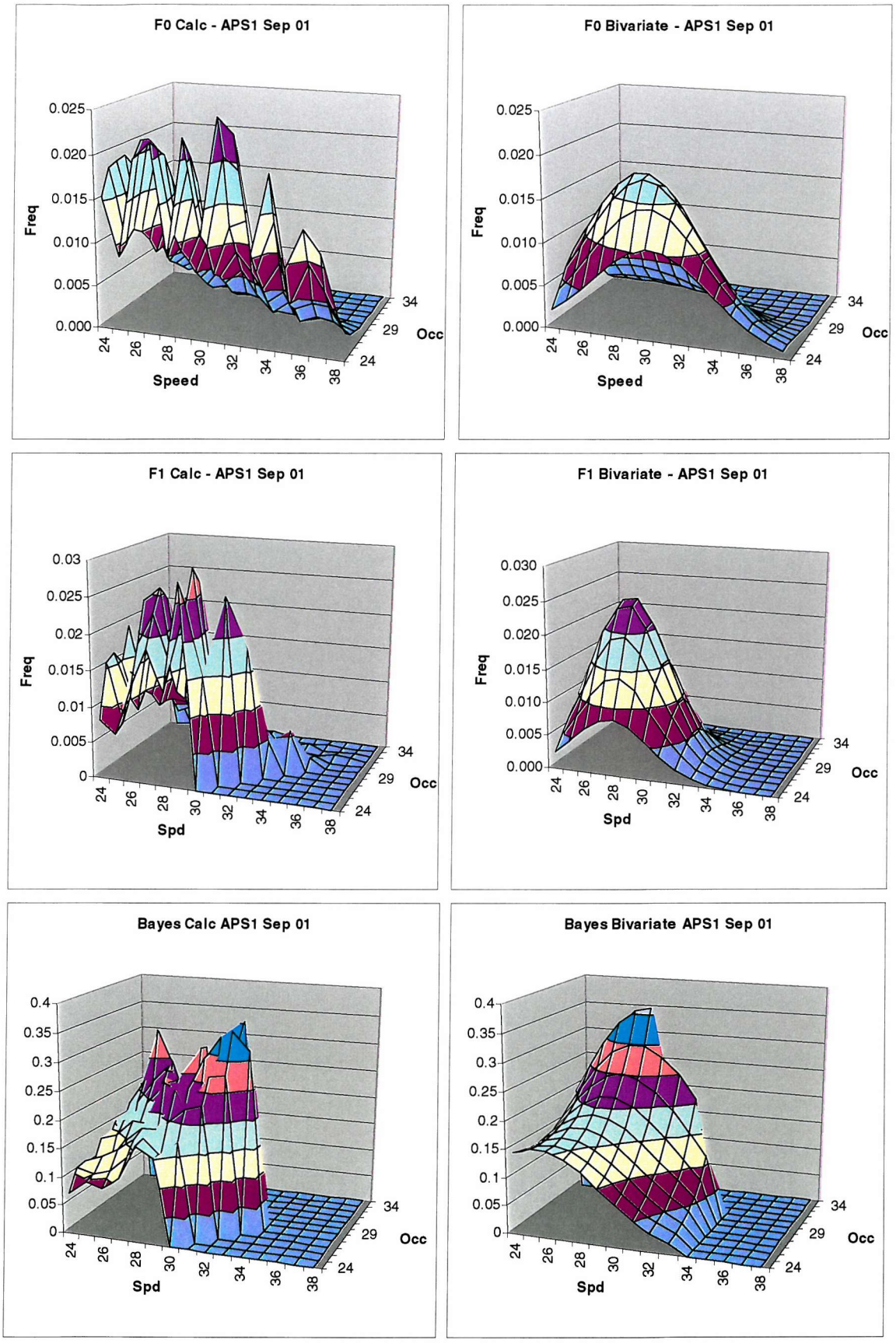
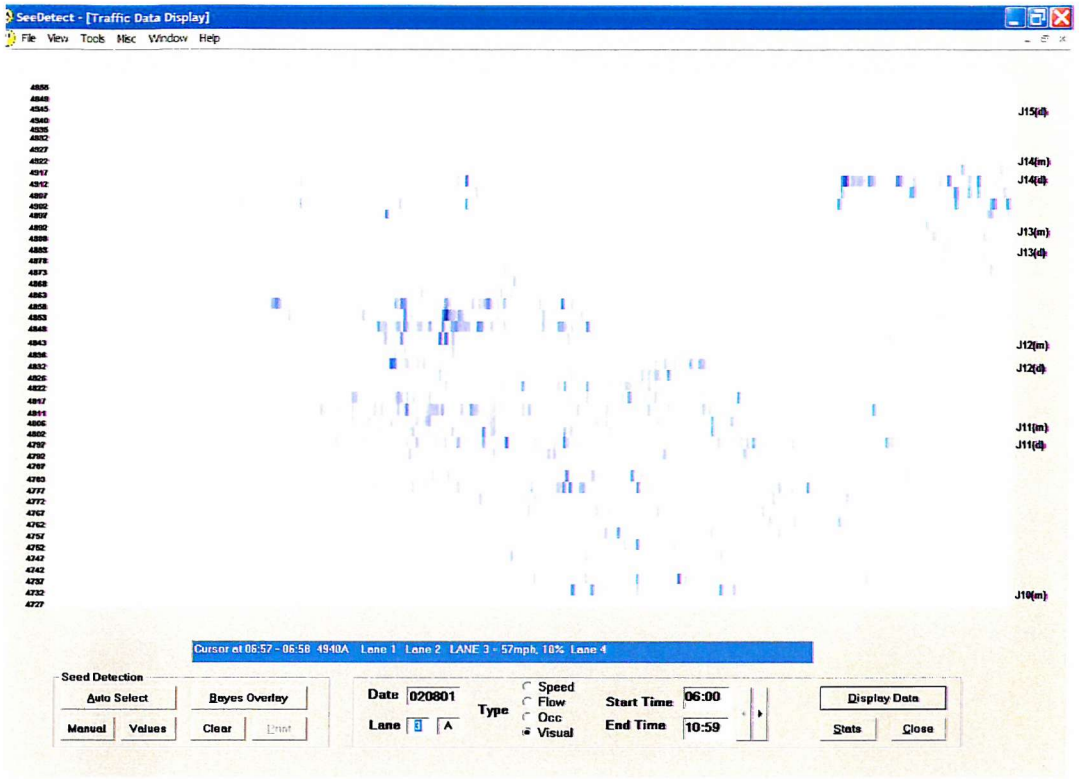


Figure E2 – Comparative plots for September 2001, showing distributions for f0, f1 and resulting posterior (using a uniform ratio for the prior function)



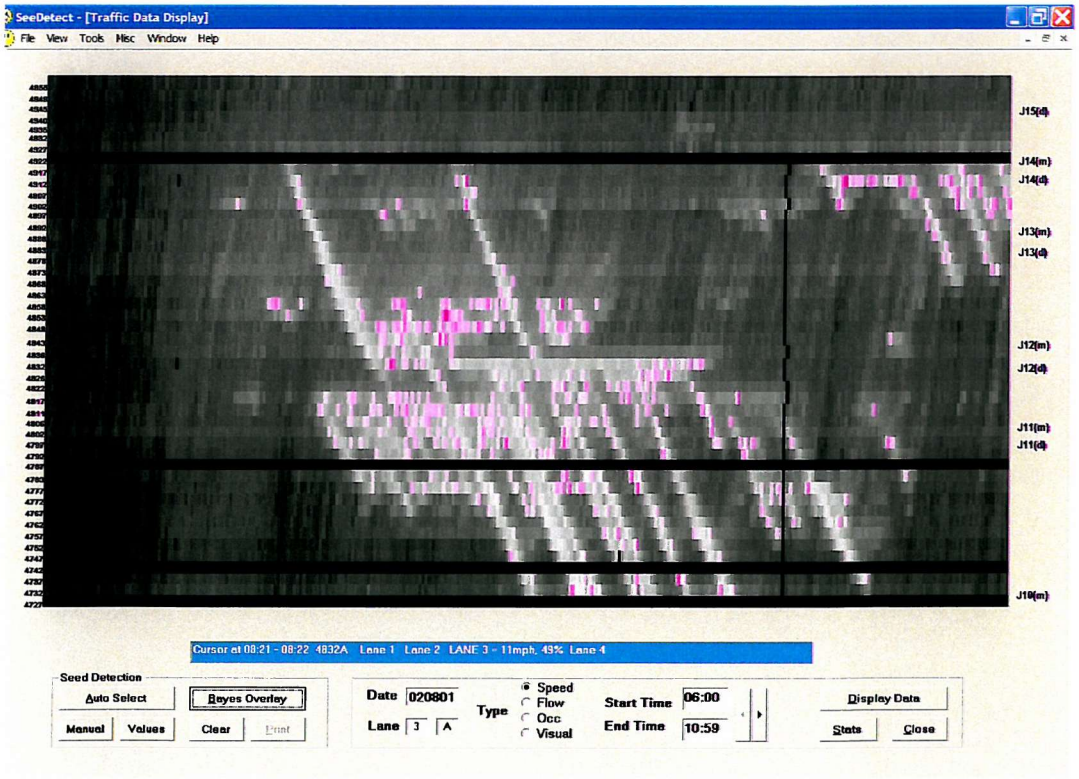
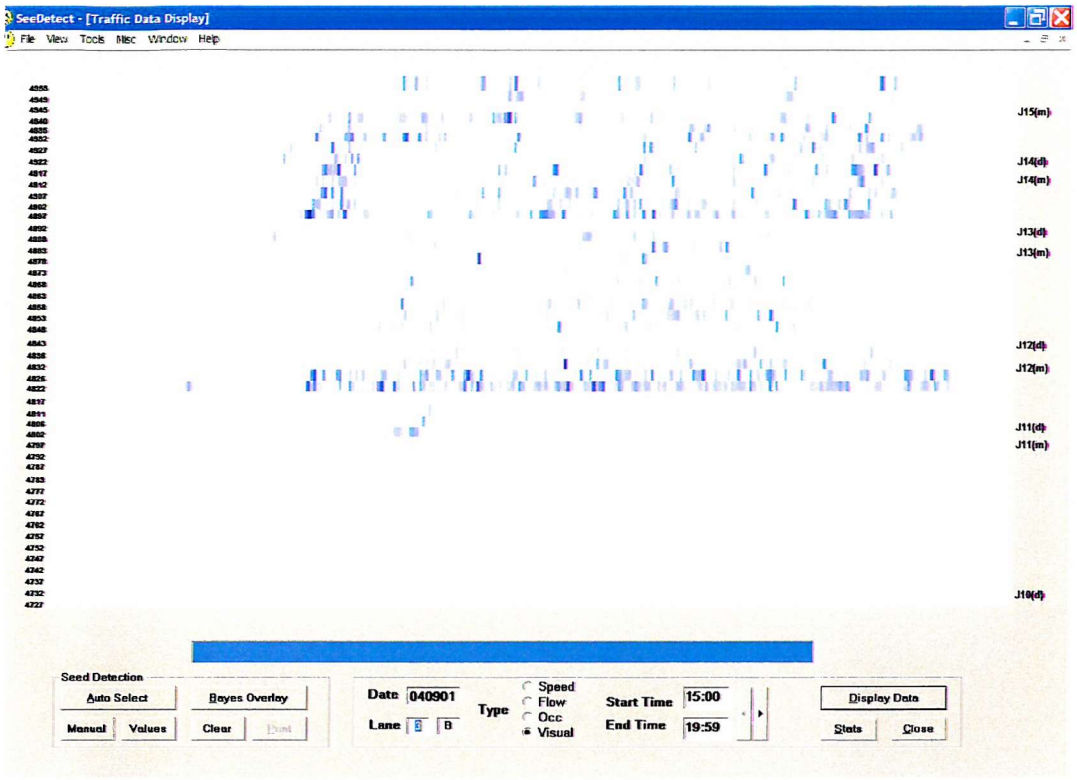


Figure E3 – Probability map for 02 August 2001 shown with and without a background of speed (grey-scale with shockwaves shown in white). Clustering around seed points is evident.



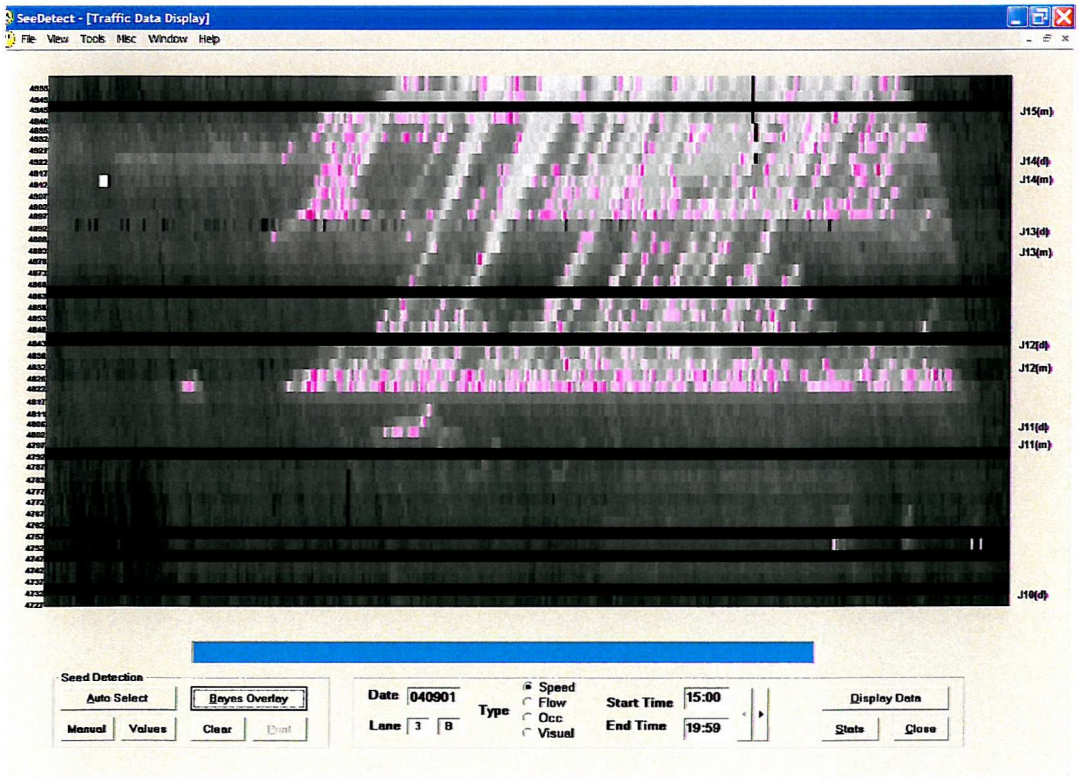


Figure E4 – Probability map for 04 September 2001 shown with and without a background of speed (grey-scale with shockwaves shown in white). The method works equally well for the anticlockwise carriageway

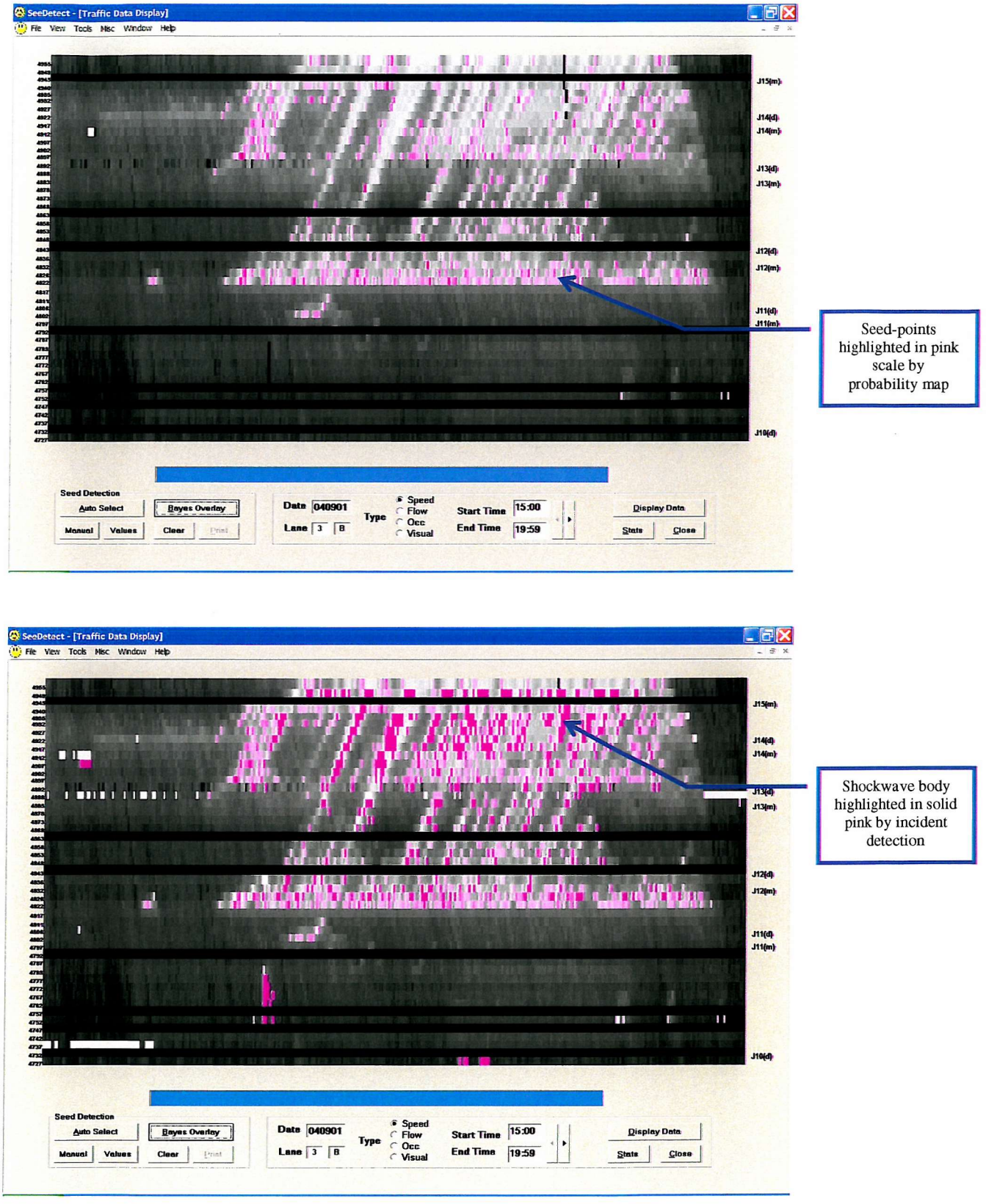


Figure E5 – Probability map for 04 September 2001 shown with and without additional information from the incident detection algorithm (California)

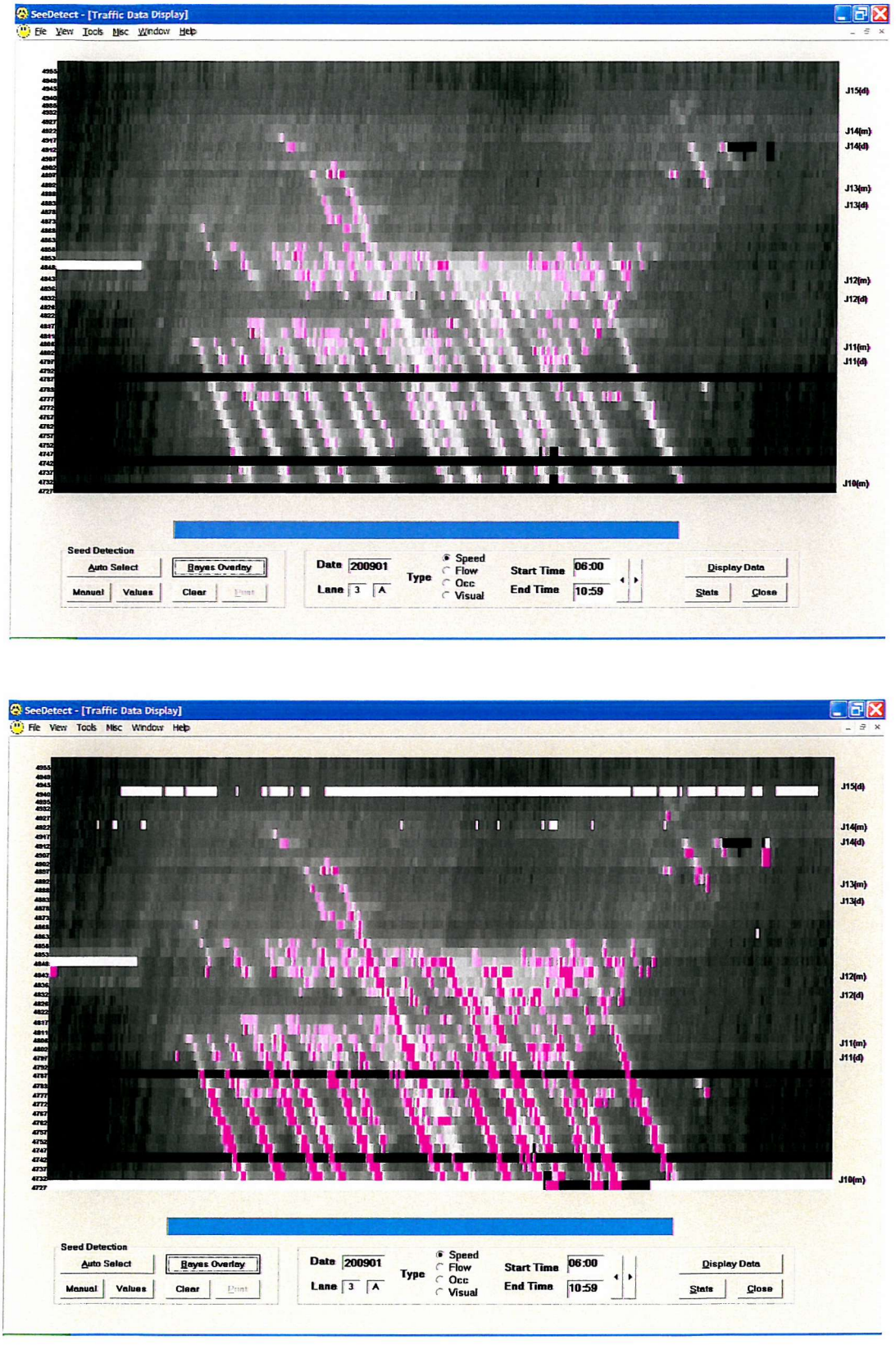


Figure E6 – Probability map for 20 September 2001 shown with and without additional information from the incident detection algorithm (California)

APPENDIX F
KEY SOFTWARE LISTINGS

F1 SEED DETECTION ALGORITHM

```
'*************************************************************************
'* This is where the rules reside, see previous algorithm for controls *
'*************************************************************************
'
Private Function blnExecuteRule(ByVal intRule As Integer, ByVal intLoopNo As Integer,
ByVal intTime As Integer, ByVal intSpeed As Integer, ByVal intFlow As Integer, ByVal
intOccupancy As Integer, ByVal intShockwave As Integer) As Boolean
  Dim intNextSpeed As Integer
  Dim intNextOccupancy As Integer
  Dim intCounter As Integer

  'Assume the rule doesn't apply
  blnExecuteRule = False

  Select Case intRule ' Called from the previous routine CmnGo

    Case 1  'Rule 1: Range for occupancy (user defined)
        If (intOccupancy >= Val(txtMinOcc.Text)) And (intOccupancy <=
Val(txtMaxOcc.Text)) Then
            blnExecuteRule = True
        Else
            blnExecuteRule = False
        End If

    Case 2  'Rule 2: Range for speed (user defined)
        If (intSpeed >= Val(txtMinVel.Text)) And (intSpeed <= Val(txtMaxVel.Text)) Then
            blnExecuteRule = True
        Else
            blnExecuteRule = False
        End If

    Case 3 'Rule 3: Time Series (Speed Falling and Occupancy Rising)
        intNextSpeed = pclsLoopdata(intLoopNo).intSpeed((intTime - 1),
Val(txtLaneNo.Text))
        intNextOccupancy = pclsLoopdata(intLoopNo).intOccupancy((intTime - 1),
Val(txtLaneNo.Text))
        If (intOccupancy > intNextOccupancy) And (intSpeed < intNextSpeed) Then
            blnExecuteRule = True
        Else
            blnExecuteRule = False 'must reset if there is a problem
        End If

    Case 4 'Rule 4: Check the difference in values (indicates what type of traffic)
        If (intSpeed - intOccupancy) >= Val(txtFactor1.Text) And (intSpeed -
intOccupancy) <= Val(txtFactor2.Text) Then
            blnExecuteRule = True
        Else
            blnExecuteRule = False
        End If

    Case 5 'Rule 5: Checks for a reversal in time (to edge of shockwaves)
        intNextSpeed = pclsLoopdata(intLoopNo).intSpeed((intTime + 1),
Val(txtLaneNo.Text))
        intNextOccupancy = pclsLoopdata(intLoopNo).intOccupancy((intTime + 1),
Val(txtLaneNo.Text))
        If (intNextSpeed - intNextOccupancy) >= Val(txtNextDelta.Text) Then
            blnExecuteRule = True
        Else
            blnExecuteRule = False
        End If

  End Select
End Function
```

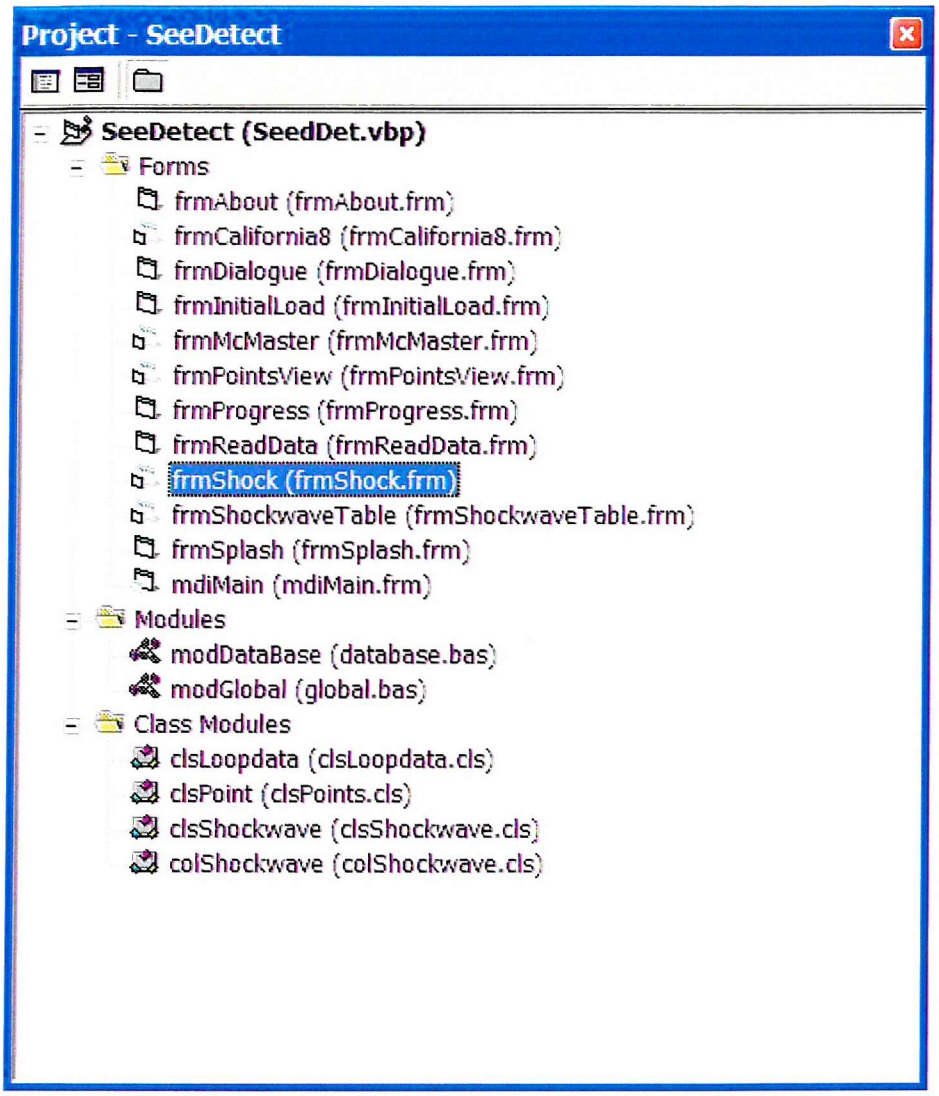
F2 CALIFORNIA #8 ALGORITHM

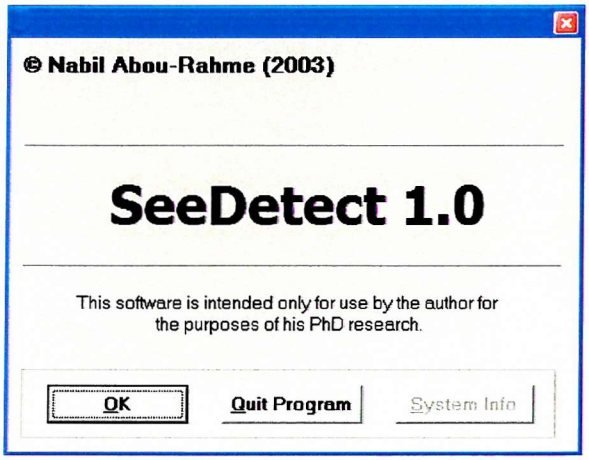
```
        LOOP for all available pairs of detectors

        consecutive_failure = 0

        detector_state = CLEAR

        LOOP for all available minutes of data

        difference = upstream_occupancy(t) - downstream_occupancy(t)

        relative_difference = upstream_occupancy(t) - downstream_occupancy(t)
                              ---------------------------------------------
                                        upstream_occupancy(t)

        relative_variation = downstream_occupancy(t-2) - downstream_occupancy(t)
                             -----------------------------------------------
                                      downstream_occupancy(t-2)

        IF  detector_state = CLEAR
[check for a new compression wave]
                IF  downstream_occupancy(t) >= compression_threshold
                        AND relative_variation < relative_variation_threshold THEN
                                detector_state = COMPRESSION_1
[check for a new incident]
                ELSEIF difference >= difference_threshold
                                AND relative_difference >= relative_difference_threshold
                                AND downstream_occupancy(t) < incident_threshold  THEN
                        detector_state = ALARM_TENTATIVE
                ENDIF
[check for ongoing compression wave]
        ELSEIF  detector_state = COMPRESSION_n THEN
                IF detector_state = COMPRESSION_5 THEN
                        detector_state = CLEAR
                ELSE
[increment the time count of the compression wave]
                        detector_state = COMPRESSION_n+1
                ENDIF
[ongoing incident]
        ELSE
                IF relative_difference >= relative_difference_threshold THEN
                        IF detector_state = ALARM_TENTATIVE
                                detector_state = ALARM_CONFIRMED
                        ELSE
                                detector_state = ALARM_CONTINUING
                        ENDIF
[check if the incident is a compression wave]
                ELSEIF relative_variation < relative_variation_threshold
                        AND downstream_occupancy >= compression_threshold
                        AND detector_state = ALARM_TENTATIVE THEN
                                detector_state = COMPRESSION_1
                ELSE
[clear incident]
                        detector_state = CLEAR
                ENDIF
        ENDIF

        IF detector_state = CLEAR THEN
                consecutive_failure = 0
        ELSE
                consecutive_failure = consecutive_failure + 1
        ENDIF

        IF consecutive_failure >= persistence_threshold THEN
                RAISE_ALARM
        ENDIF
```

F3 MCMASTER ALGORITHM

```
        LOOP for all available pairs of detectors

        consecutive_failure = 0

        detector_state = CLEAR

        LOOP for all available minutes of data

        IF occupancy(t) >= occupancy_threshold THEN
[this is area 3]
                detector_state = CONGESTION
[where q(x) is the least squares optimized quadratic fit calibrated at this detector]
        ELSEIF flow(t) < q(occupancy(t)) - standard_deviation* constant_scale THEN
[this is area2]
                detector_state = INCIDENT
        ELSE
[must be area1]
                detector_state = CLEAR
        ENDIF

        IF detector_state = CLEAR THEN
                consecutive_failure = 0
        ELSE
                consecutive_failure = consecutive_failure + 1
        ENDIF

        IF consecutive_failure >= persistence_threshold THEN
                RAISE_ALARM
        ENDIF
```

F5 SEED DETECTOR SOFTWARE – SCREENSHOTS





F6 ACKNOWLEDGEMENTS

SeeDetect© is a bespoke software research tool developed by Nabil Abou-Rahme to assist with PhD research into the origin of shockwaves. Certain features within this software might eventually form the basis for a larger development in traffic software, but this is outside the scope of the present study.

MTV© is a graphical research tool, initially the brainchild of Peter Still at TRL Limited, developed in association with Nabil Abou-Rahme to run with MS Windows. This package can be purchased directly from TRL (softwarebureau@trl.co.uk or www.trl.co.uk).

APPENDIX G
GLOSSARY OF TERMS

'A' Carriageway	In general terms the 'A' and 'B' labels distinguish each direction of the main carriageway from the slip roads. In context, 'A' Carriageway means the *clockwise* carriageway of the M25 (northbound on the Controlled Section).
AADT	Average Annual Daily Traffic, 24-hour flows averaged over the whole year.
AAWT	Average Annual Weekday Traffic, 24-hour flows averaged over weekdays only for the whole year.
Accident	An incident resulting in damage or injury.
Alert	An event in the MIDAS system, which signifies that a certain condition, has occurred, e.g. passing a flow threshold.
Algorithm	A method embodied in a software program for setting variable speed limit or other signs in response to flow or other alerts generated by the MIDAS system.
'B' Carriageway	In general terms the 'A' and 'B' labels distinguish each direction of the main carriageway from the slip roads. In context, 'B' Carriageway means the *anticlockwise* carriageway of the M25 (southbound on the Controlled Section).
Bayesian	A branch of inferential statistics associated with Thomas Bayes. Using the principles of conditional probability, an estimated prior distribution is then combined with empirical observations to derive an improved distribution known as the posterior.
Bottleneck	A point of restriction along the carriageway, usually described in terms of reduction in flow or capacity. Sometimes caused by a change in geometry (for example, a lane-drop), and sometimes by driver behaviour (for example, a change in headway distribution).

Capacity	A measure of the maximum possible throughput of a road section, normally expressed as the number of vehicles which can be handled in a given time period. It is not well defined on motorways, due to flow-breakdown.
Car-following	The behaviour of vehicles relative to each other (often on a longitudinal plane) described in mathematical terms using distance, speed and acceleration. Many car-following algorithms contain an implicit speed-headway relationship.
Compliance	Conformity with speed limits.
Continuity equation	Mathematical analysis found in most branches of science and engineering, based upon the first law of thermodynamics – namely the conservation of energy.
Controlled Motorways	A motorway environment comprising gantry-mounted variable speed limits driven by information gathered from loop detectors at regular intervals. A pilot scheme is in place between Junction 10 and Junction 16 on both carriageways of the M25.
Diverge	Common reference to the slip road which takes traffic off the main carriageway and onto the local road network or another motorway
Downstream	Further ahead on a motorway link or moving in the direction of travel.
EMS	Enhanced Message Signs, used to provide additional information in conjunction with the variable speed limits
Flow-breakdown	A transition, in which heavy traffic flow ceases to be smooth. Associated with a sharp fall in speed and a temporary reduction in effective road capacity.
Headway	The separation between the front of one vehicle and the front of the next vehicle. Headway is given in terms of time (often seconds) or distance (often metres).
Highways Agency	An executive agency of the UK Government's Department for Transport. The Highways Agency has primary responsibility for the UK Motorway and Trunk Road Network.

HIOCC	HIgh OCCupancy incident detection algorithm implemented on the Controlled Section during October 1997 and used to protect queues from faster traffic upstream.
Hydrodynamic	Analysis which has to do with the movement of fluids, and which is often applied to traffic to derive analogous behaviour.
Incident detection	Algorithms to analyse detector data for patterns associated with traffic build-up around an incident. More recent applications include tracking the tail of a moving queue (or shockwave).
IVD	Individual Vehicle Data. Raw traffic data collected for selected sites and periods by linking portable PCs to MIDAS substation units.
Loop detectors	A loop of wire buried in the road surface, which produces an electrical response when a large metal object, such as a vehicle, passes close overhead.
Merge	Common reference to the slip road which brings traffic onto the motorway. The 'merge-point' is where traffic actually joins the main carriageway.
MIDAS	Motorway Incident Detection and Automatic Signalling system, which processes traffic information collected through loops and determines the most appropriate speed limit to be displayed on local gantries.
Mode	The value of a variable most frequently observed.
Motorway network	Comprising all M-Class roads throughout England, for which the Highways Agency is responsible (see also 'Trunk road network').
MTV	Motorway Traffic Viewer – a suite of graphical analysis programs (for which the author was a co-developer at TRL Limited). Used to examine MIDAS data.
Nearside	A label to describe the furthest left on a carriageway or the passenger's side of the vehicle (for UK roads). The nearside lane is often termed 'the slow lane'.
Occupancy	A proxy measurement for headway. Occupancy for a loop detector is the % time for which that loop is covered (or occupied).

Offside	A label to describe the furthest right on a carriageway or the drivers' side of the vehicle (for UK roads). The offside lane is nearest to the central reservation.
Peak 15-minute	The highest value, usually of throughput, measured in any 15-minute period.
Peak hourly	The highest value, usually of throughput, measured in any one-hour period.
Performance Indicator	A numerical measure of traffic behaviour on the motorway which has implications for assessing the performance of the Controlled Motorway system.
Platoon	In context this refers to a cluster of vehicles travelling together with similar speeds and spaced by relatively short headway.
Prior distribution	Represents initial belief about the distribution of a random variable, which can then be combined with empirical data to produce an improved estimate (see also 'Posterior distribution').
Probabilistic model	A model that includes probability-based calculations to derive its outputs.
Posterior distribution	A probability distribution associated with inferential statistics. Derived by combining the initial belief with empirical data (see also 'Prior distribution').
Ramp metering	A method for controlling access from the slip road in order to maximise flow on the main carriageway.
Rubbernecking	Drivers slowing to look at an event or incident on the *opposite* carriageway will cause congestion on their own carriageway also.
Seed-point	A point on the motorway where flow-breakdown first occurs. More specifically, a point to which the origin of a shockwave can be traced and from which that shockwave appears to be propagating.
Shockwave	(a) A wave that moves through a fluid faster than a change can normally be propagated. (b) A wave of flow-breakdown which propagates upstream through a column of moving traffic. A flow-breakdown shockwave is characterised by a constant speed of propagation, and by an

increasingly large and abrupt drop in the speed of traffic as it passes, which can lead eventually to stop-start driving.

(c) Commonly used to describe the boundary between the two regions rather than the high-density region itself.

Significance The probability that an observation occurred by chance, assuming some underlying probability structure (e.g. a Normal distribution). Significance is usually expressed as a percentage, and the smaller it is the better. 5% or less is usually taken to indicate a significant observation or result.

Significant Unlikely to have occurred by chance.

Site numbering A marker-post with intervals of 100m. For example, on the M25, site 4902A is at kilometre position 90.2 midway between Junctions 13 and 14 on the A-Carriageway. Site numbers run from 4727 at the south end of J10, to 4955 at the north end of J15. The number 4 identifies the M25 motorway in this case.

Stop-start Driving conditions in which some vehicles temporarily come to a halt.

Swooping Drivers moving late from the offside lanes onto a diverge lane, typically crossing more than one lane in a single manoeuvre, are said to be 'swooping' onto the junction.

Throughput The amount of traffic passing through a given road section in a certain period of time.

TRL The Transport Research Laboratory, formerly an executive agency for the UK Department for Transport, transferred to the private sector as a non-profit organisation in 1996. Now known as TRL Limited.

Trunk road network Comprising primary A-Class roads throughout England, for which the Highways Agency is responsible (see also 'Motorway network').

Tuning The process of adjusting parameters to produce a desired change or improvement in system performance, usually in the light of monitoring results.

Typical day	A weekday representative of the majority of such days in a particular year – excluding holidays, bad weather and incidents.
Upstream	Further behind on a motorway link or moving against the direction of travel.
Utilisation	Usually the proportion of traffic using a given lane.
Variable speed limit	Gantry mounted mandatory speed limit signs, displayed to motorists on the Controlled Motorway.
Vehicle Hour Delay (VHD)	A summation of the delay experienced by drivers who are not able to travel at a speed of 50 mph. Used as a means of evaluating congestion.
Vehicle trajectories	The path taken by a vehicle, described in terms of distance and time.
Vehicular composition	Usually a reference to the fact that traffic flow comprises more than one class of vehicle (for example, cars and HGV).

APPENDIX H

THE BOUNDARY SPEED EQUATION

In Chapter 2 of this thesis, the boundary speed equation identified by Lighthill and Whitham (1955) was presented as a key characteristic of shockwave propagation. The equation for this boundary speed (u) is stated in terms of density and speed as follows:

$$u = \frac{(k_1 v_1 - k_2 v_2)}{(k_1 - k_2)}$$

assuming that $q = kv$ holds true for the specific traffic data being observed. Since $k_1 < k_2$ and $v_1 > v_2$ the value of u is often negative suggesting backward propagation against the direction of travel.

During the development of this thesis a question arose about whether the boundary speed could be verified using the available MIDAS data. The following reasons formed part of the initial objection to making this application:

1. The boundary speed equation is stated in terms of speed and density for the two regions. The detector data available through MIDAS provides speed and flow averaged over one-minute intervals and at a resolution of around 500m. However, the 'density' measurement is a proxy called 'occupancy' which describes the percentage time that the loop is active during a particular minute. Whilst it is possible to show some mapping from this value to the 'vehicles per unit length' value it is not straightforward. This compels movement in the direction of first accepting that q=kv must hold for this data set, and second, making the appropriate substitution into the boundary speed equation so that it is stated in terms of speed and flow only.

2. The temporal and spatial resolution of the data causes some difficulty with identifying the exact position of the boundary. Over a number of loops it is possible to measure the gradient of the shockwave and to conclude that the speed of propagation is around 18–23kph. But

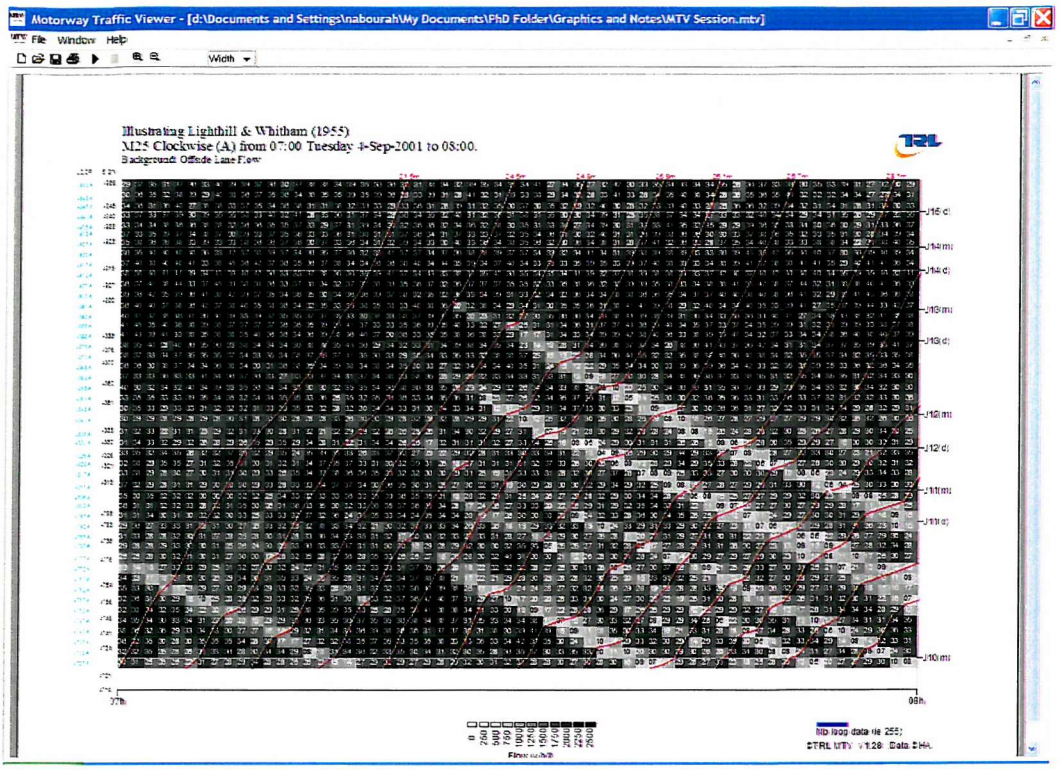
demonstrating that through an equation which describes the boundary is not so straightforward. It is comparable to the difficulty with identifying a relationship between traffic characteristics at the upstream detector and the identification of a seed point – a lot can change in 500m (see existing text in Section 5.3 and Section 6.3.4).

The SeeDetect© software was modified in order to address this question in numerical terms. The path of a virtual vehicle was mapped through the traffic data (the closest approximation achievable with this data for a group of vehicles passing through a number of shockwave boundaries). A good implementation of these algorithms has been developed by TRL Limited for use with the MTV Software. A minor modification to the algorithm allowed a comparison between speed and flow prevalent at the location of the virtual vehicle over time.

The boundary speed equation had to be restated in terms of v (speed in mph) and q (flow in vehicles per minute). In practice, application of the equation yielded meaningful results only at the point when a rapid drop in speed took place (in other words, when a boundary was clearly present). At other times the value of u was set to zero in order to remove that background noise.

As expected, the results were somewhat disappointing. Figure H1 shows the virtual vehicle trajectories and Figure H2 shows the treatment of the values obtained. Figure H3 provides a direct comparison between the vehicle trajectory and the calculated boundary speed, similar to that presented in Figure H1 and Figure H2.

The primary difficulty for application of the equation to this data was the spatial and temporal resolution which led to the need for an approximation of vehicle trajectory and the large variation in the results (despite empirical measurements of shockwave gradient yielding such consistent results). This exercise should be repeated when more closely-spaced detectors become available (for example, on the M42 ATM Pilot).



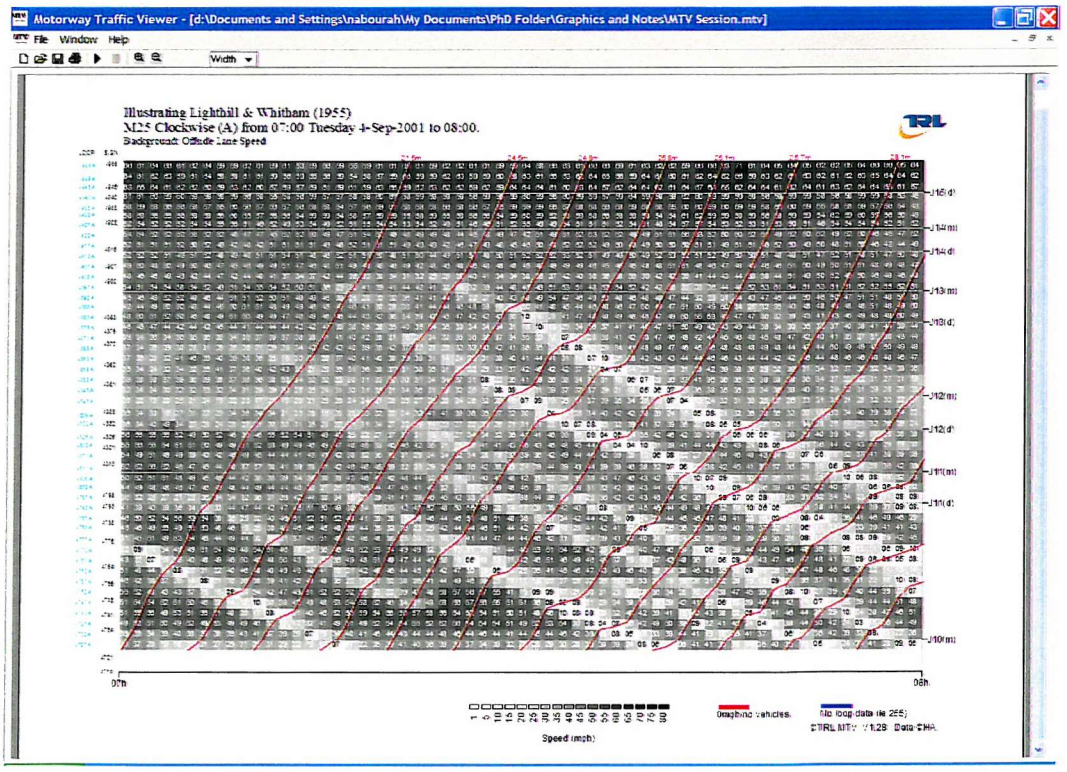


Figure H1 – The MTV Plots above illustrate the path taken by virtual vehicles which may represent the experience of a real driver on that day. The trajectories in each plot are identical but the background displays speed and flow respectively for 04 September 2001.

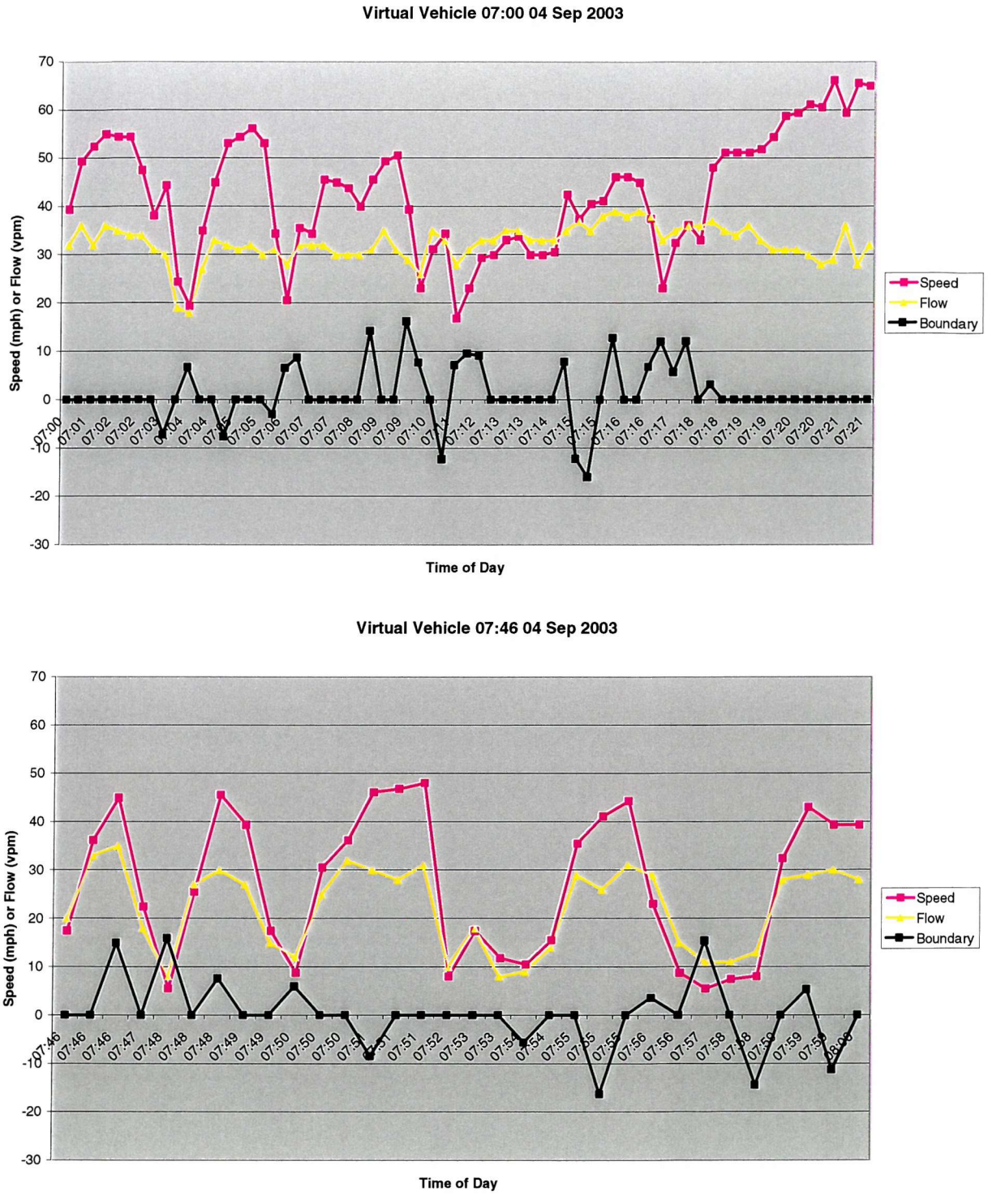


Figure H2 – Speed and flow profiles over the journey are plotted for vehicles commencing from M25 J10 at 07:00 and 07:46 respectively. The value of the 'boundary' calculation varies excessively so only those values falling within the absolute range of 0–20mph are included here. A rapid fall in speed indicates when the vehicle has crossed the boundary. On some occasions the boundary speed appears to be in the right range (12–15mph) but on other occasions no such pattern is evident.

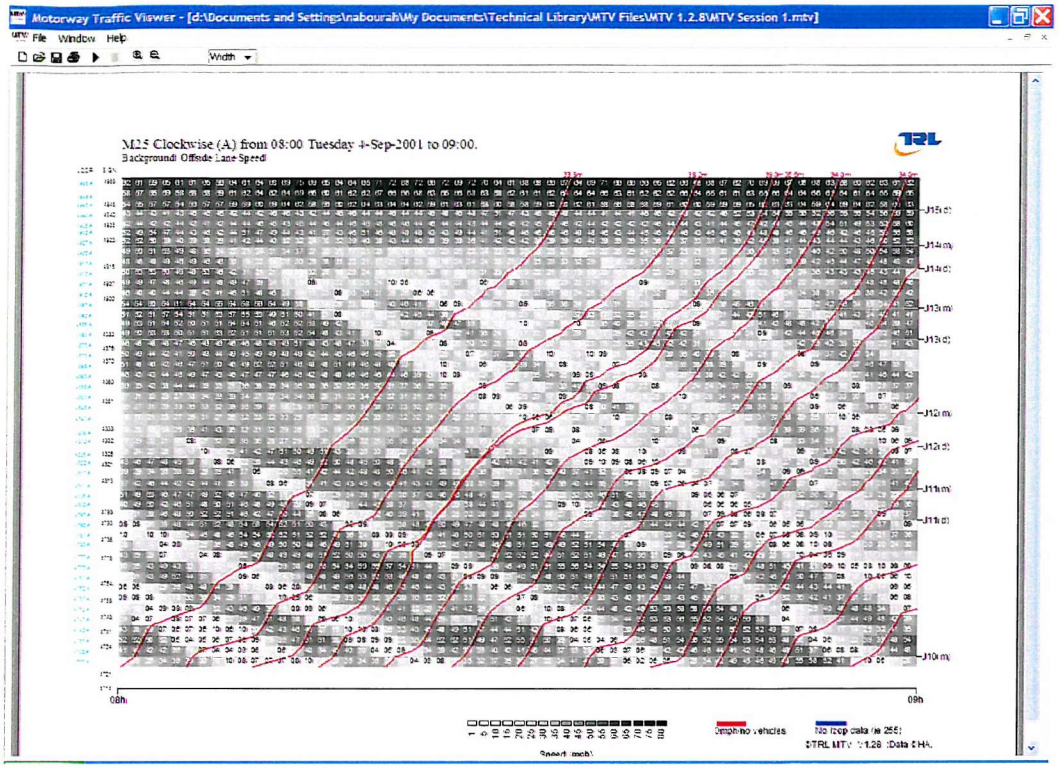


Figure H3 – The top plot shows virtual vehicle trajectories for the hour between 08:00 and 09:00 on 04 September 2001. The bottom plot shows the speed and flow values extracted from the first of these trajectories, commencing from J10 at 08:00. Fluctuations in the boundary speed seem more consistent in the second half of the journey from around 08:20 onwards.