Modelling Congestion Caused by Abnormal Loads

by Nicholas Taylor

PPR196
PPAD 9/33/123

PUBLISHED PROJECT REPORT
MODELLING CONGESTION CAUSED BY ABNORMAL LOADS

Issue 2

by Nicholas Taylor (TRL Limited)

Prepared for: Contract: PPAD 9/33/123
Evaluating Congestion Caused by Abnormal Loads
Client: AIL Unit, Highways Agency
(Mr. Andrew Cook)

Copyright TRL Limited September 2007

This report has been prepared for the AIL Unit of the Highways Agency. The views expressed are those of the author(s) and not necessarily those of the Highways Agency or the Department for Transport.

Published Project Reports are written primarily for the Customer rather than for a general audience and are published with the Customer’s approval.
This report has been produced by TRL Limited, under a contract placed by the Highways Agency. Any views expressed are not necessarily those of the Highways Agency.

TRL is committed to optimising energy efficiency, reducing waste and promoting recycling and re-use. In support of these environmental goals, this report has been printed on recycled paper, comprising 100% post-consumer waste, manufactured using a TCF (totally chlorine free) process.

History:
Issue 1: December 2006
Issue 2: September 2007 – revision to Section E.7
CONTENTS

Executive Summary iii

1 Introduction 1

2 Principles of the model 1

3 Implementation in spreadsheet model 2

4 Calibration and other practical considerations 3

5 Additional contributions to economic cost 4

6 Conclusion and results 5

7 Acknowledgements 6

8 References 6

APPENDIX A - TYPES OF QUEUE AND DELAY MODEL 7

A.1 Dunn’s formula and delay estimation from direct measurements 7
A.2 ‘Vertical’ versus ‘Horizontal’ queue models 7
A.2 Theoretical models of queuing and capacity 9
A.3 Comparison of models 10

APPENDIX B – TECHNICAL DESCRIPTION OF THE MODEL 11

B.1 Speed/flow/density relationships 11
B.2 Speed and flow in the vicinity of a load 12
B.3 Queue caused by the load 15
B.4 Interpretation of queue in conventional demand and capacity terms 16
B.5 Total delay caused by the load 16
B.6 Estimating average delay per vehicle 18
B.7 Linking of queues between successive route sections 18
B.8 Knock-on effect of released queues 19
B.9 Modelling the effect of diversion 20
B.10 Summary and calibration of model 21
Executive Summary

M.1 Congestion on Britain’s roads is an increasing problem and in recent years traffic volumes have increased to such an extent that parts of the network are often operating close to capacity. The Department for Transport (DfT) and the Highways Agency (HA) are committed to reducing traffic congestion by better management of the road network. As part of this commitment alternative modes of transport are being considered.

M.2 The movement of abnormal loads through the network causes additional delays which could be reduced by a better understanding of how delays are related to the type of load and the conditions under which they are moved. TRL has been asked to provide the Highways Agency with a better understanding of how different types of abnormal load, travelling on different types of road and under different conditions, influence traffic on routes along which they travel and in surrounding areas. This will assist the Abnormal Loads team in the Highways Agency in deciding through a ‘test reasonableness’ whether to permit the largest and heaviest abnormal loads to travel by road, and to promote alternative method of transporting abnormal loads, in particular by water.

M.3 It is difficult to measure the effects of abnormal loads directly, especially on general purpose roads which are not instrumented. Direct observations, made either from the roadside or travelling with the load, can provide confirmation of estimates, but are too expensive for more than a sample of loads. Therefore a model is needed to estimate traffic conditions and delays given the characteristics of the load and the road sections over which it travels.

M.4 The model's results for congestion costs will contribute to the Highways Agency’s assessment of abnormal load applications. Other externality costs are identified and estimated, but their detailed evaluation and calculation for abnormal loads is a separate issue. The estimated externality cost per mile can be input into the model to form part of the overall costs, provided there is an acceptance that it is based on the most relevant information currently available.

M.5 The model is implemented in a spreadsheet in which a load’s route can be analysed section by section to describe the progress and effect of the load and to provide an estimate of the cost-per-mile of the delay which it causes. This isolates the components of delay caused by an abnormal load, expressing them in terms of basic traffic flow parameters. All types of road from multi-lane dual to minor single carriageway are covered. Blocking of both directions of a single carriageway can be represented, and by suitable definition of sections queuing relieved periodically by pulling the load off the road or other traffic management measures. Account is taken of average traffic speeds and densities in free flowing and queuing traffic as well as the numbers of vehicles queuing.

M.6 The main report describes the principles and general findings. Appendices A-F address the different types of queue and delay model, the detailed description of the present model, approaches to estimating demands and capacities, making measurements while travelling with a convoy, discussion of monitoring results including a real case example, economic cost and other impacts discussed below.

M.7 Calibrations or assumptions needed are:

(1) Description of the route, road type (Dual or Single, no. of lanes), section capacities, demand flows (depending on time of day), proportion of flow which continues from one section to the next, and any bottlenecks downstream which may cause queuing of heavy discharge flows;

(2) Uncongested speed/flow relationships normally applying to each road section, and to the ‘channel’ or remaining lanes available for passing the load, if any;

(3) The time and distance headway parameters governing the car-following model, unless inferred from section capacity in which case they should be confirmed as reasonable;

(4) The length of the Zone of Influence, or the length of road affected by the load and its escorts.
M.8 The costing part of the spreadsheet includes a sum of the distances and costs on all the sections and the corresponding average cost-per-mile for the whole move. By default a single economic cost/veh-h figure is used, though this can be set individually for each road section if desired. In principle, this can vary by day of the week and time of day, according to traffic composition. Up-to-date figures are available from WebTAG (http://www.webtag.org.uk). However, the practical variation is quite small and an average value can safely be assumed in most cases.

M.9 Within the model working, several types of delay are distinguished and estimated separately, depending on the setting of certain extra switches or factors. These are combined to give the total aggregate delay to be costed.

- Delay in a queue following the load
- Delay resulting from slowing down to pass the load
- Delay to traffic in the opposite direction, if the road has to be closed
- Delay caused by a heavy queue outflow hitting a bottleneck downstream
- Variation in delay according to whether traffic follows the load from one section onto the next

M.10 The primary purpose of this report and model is to assess congestion costs. However, other possible impacts or externalities include:

- Accidents
- Noise and vibration
- Road and infrastructure damage
- Emissions, resulting in either Pollution or Climate Change
- Other environmental impact
- Interference with movement

M.11 Given the variables collected for or estimated by the model, and appropriate relationships, these impacts could in principle be estimated as follows:

<table>
<thead>
<tr>
<th>Impact</th>
<th>Depends on (X journey time or mileage as appropriate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accident risk</td>
<td>Queues, difference in speed between load and passing traffic</td>
</tr>
<tr>
<td>Noise and vibration</td>
<td>Load speed, engine effort related to weight, road standard</td>
</tr>
<tr>
<td>Road wear and infrastructure</td>
<td>Axle loading, axles, load speed</td>
</tr>
<tr>
<td>Emissions</td>
<td>Load speed, weight, delay to other traffic</td>
</tr>
<tr>
<td>Environmental impact</td>
<td>Road type</td>
</tr>
<tr>
<td>Interference with movement</td>
<td>Delays, diversion, road closures</td>
</tr>
</tbody>
</table>

M.12 Some published figures for general freight are used to estimate an overall cost contribution of about £3/mile, in addition to comparisons with water and rail (see Appendix I to the main modelling report). However, there is no evidence that these figures are appropriate to Abnormal Loads, nor is there any evidence on the safety impact of queues caused by Abnormal Loads.

M.13 Diversion is a possible impact, sensitive to the nature of the surrounding network. Methods of estimating diversion caused by long-term obstructions like road works, such as QUADRO, may not be appropriate to the temporary presence of an AIL, even if warnings are broadcast. Other externalities listed in section 2.4.2 of the Literature Review, such as the cost of providing AIL facilities, are outside the brief of this case-based approach.

M.14 Results of comparing the model with direct observation and MIDAS analysis on the M6 are encouraging. Several other cases have been modelled. A separate sensitivity analysis could provide an insight into the way delay costs can vary depending on the day or time of a move, being particularly sensitive to traffic volumes close to road capacity. Perennial traffic growth can likewise be taken into account by applying growth factors to estimated traffic volumes.
1 Introduction

1.1 Congestion on Britain’s roads is an increasing problem and in recent years traffic volumes have increased to such an extent that parts of the network are often operating close to capacity. The Department for Transport (DfT) and the Highways Agency (HA) are committed to reducing traffic congestion by better management of the road network. As part of this commitment alternative modes of transport are being considered.

1.2 The movement of abnormal loads through the network causes additional delays which could be reduced by a better understanding of how delays are related to the type of load and the conditions under which they are moved. TRL has been asked to provide the Highways Agency with a better understanding of how different types of abnormal load, travelling on different types of road and under different conditions, influence traffic on routes along which they travel and in surrounding areas. This will assist the Abnormal Loads team in the Highways Agency in deciding through a ‘test of reasonableness’ whether to permit the largest and heaviest abnormal loads to travel by road, and to promote alternative method of transporting abnormal loads, in particular by water.

1.3 It is difficult to measure the effects of abnormal loads directly, especially on general purpose roads which are not instrumented. Direct observations, made either from the roadside or travelling with the load, can provide confirmation of estimates, but are too expensive for more than a sample of loads. Therefore a model is needed to estimate traffic conditions and delays given the characteristics of the load and the road sections over which it travels.

1.4 The model’s results for congestion costs will contribute to the Highways Agency’s assessment of abnormal load applications. Other externality costs are identified and estimated, but their detailed evaluation and calculation for abnormal loads is a separate issue. The estimated externality cost per mile can be input into the model to form part of the overall costs, provided there is an acceptance that it is based on the most relevant information currently available.

1.5 The model is implemented in a spreadsheet in which a load’s route can be analysed section by section to describe the progress and effect of the load and to provide an estimate of the cost-per-mile of the delay which it causes. This isolates the components of delay caused by an abnormal load, expressing them in terms of basic traffic flow parameters. All types of road from multi-lane dual to minor single carriageway are covered. Blocking of both directions of a single carriageway can be represented, and by suitable definition of sections queuing relieved periodically by pulling the load off the road or other traffic management measures. Account is taken of average traffic speeds and densities in free flowing and queuing traffic as well as the numbers of vehicles queuing.

1.6 The report is organised as follows: The main report describes the principles and general findings. Appendices A-E address the different types of queue and delay model, with detailed description of the present model including approaches to estimating the effect of breaks, and linking or spill-back of queues between road sections. Appendix B can be considered the ‘main’ model section. Up to and including Appendix B, figures, tables and equations are given simple numbers, except where equation numbers are of the form \{letter number\} where the letter indicates a particular context or class. Appendices F-I describe respectively approaches to estimating demands and capacities, making measurements while travelling with a convoy, monitoring results including a real case example, and economic cost and other impacts (see also Section 5 later). In Appendices C-I, figures, tables and equations are numbered as within the Appendix, eg \{letter • number\}.

2 Principles of the model

2.1 An abnormal load is treated formally as a moving blockage (see also Taylor 2004) which, depending on its nature and that of the road, may or may not allow traffic to pass in the opposite or same direction. Its route is treated formally as a series of sections with constant properties. This is a simplification because the characteristics of roads, especially single carriageways, can vary considerably as a result of narrowings, bend and junctions, and the speed of loads can be particularly sensitive to gradient. Nevertheless as many sections as required can be defined.
2.2 Because the load is moving, the way queues develop is different from that applying to fixed obstructions. The queue model takes implicit account of the average speeds and densities of traffic in different areas. As the load moves from one section to the next, it may leave a queue to discharge unimpeded on its original route, or may take with it some or all of the queuing traffic. The model must take account of the effect of queues ‘left over’ in either case. Depending on the circumstances, the time taken by a queue to disperse can be much shorter or longer than the time it took to build up.

2.3 Moving traffic is best described using speed/flow/density relationships. The present model develops certain such relationships in a simplified description involving four regions of traffic, as shown in Figure 1, in each of which traffic is assumed to be uniform.

![Figure 1. Four-region schematic of abnormal load and associated traffic](image)

2.4 The letters used to label the four regions also serve as subscripts of algebraic quantities: arrival (a), queuing or bottleneck (b), channel-past-load (c), and downstream or discharge (d), and of course the load itself (L). If it exists, the queue is essentially the dense region (b) just upstream of the abnormal load’s ‘Zone of Influence’. Its growth and decay can be described in terms of the motion of the boundaries between it and the surrounding regions, along with the so-called ‘fundamental relationship of traffic’, which states that flow (q) is the product of speed (v) and density (k).

3 Implementation in spreadsheet model

3.1 The model has been implemented in an Excel spreadsheet. Each road section is represented by a single column of formulae, which can be duplicated to as many road sections as are needed. There is a linkage between successive sections to account for all or some of the queuing traffic from one section following the load onto the next, which increases delay compared to independent sections. It is straightforward to subdivide sections to represent the load stopping to let queues pass.

3.2 The spreadsheet model consists of the following main blocks for each road section (column):

- Section descriptive text
- Carriageway data (principal direction)
- Load data (with provision for indicating lanes free/blocked, including reverse direction)
- Carriageway modelling options (including knock on effects and queue transfer between sections)
- Costing results

3.3 The model gives a single delay and cost prediction for each section based on its values of road type, demand, capacity, proportion of HGVs, section length, Zone of Influence and load speed. The section results are totalled to give the delay and delay cost for the whole route. A tabular format covering a range of possible data values (as in personal communication by P Gray and D Farley, ITEA) would require the output cells to contain a single complex formula, which is impractical.

---

1 These situations for which subscript d is used are essentially unrelated.

2 Only the relationship $q = vk$ is fundamental, because it follows from the conservation of vehicles. The parabolic type of speed/flow relationship sometimes assumed is not fundamental as it depends on further assumptions.
because of the complexity of the model. However, it is easy to set up examples with a range of demand flows, and to compare the results with those of alternative approaches (see Appendix A).

3.4 The extra carriageway data include an optional demand for reverse direction flow (defaulted to the demand in the co-moving direction), which will be required if a load forces a single carriageway road to be closed, plus switches and factors to allow for ‘knock-on’ queuing caused by the primary queue discharge flow hitting a user-defined bottleneck downstream, and for a proportion of a queue following the load on to the next section. The costing block includes a sum of the distances and costs on all the sections and the corresponding average cost-per-mile for the whole move.

3.5 Within the model working, several types of delay are distinguished and estimated separately, these being combined to give the total aggregate delay to be costed. By default a single cost per vehicle-hour figure is input, though this can be set individually for each road section if desired:

- Delay in any queue following the load
- Delay resulting from slowing down to pass the load
- Delay to traffic in the opposite direction, if the road has to be closed
- Delay caused by heavy queue outflow hitting a bottleneck downstream
- Variation in delay according to whether or not traffic follows the load from one section onto the next

3.6 The calculations are identical for all types of road from dual-carriageway motorways down to single carriageway minor roads. The different characteristics of roads are reflected in their speed/flow parameters and capacities. However, the model employs simple switches or factors to activate certain features. These include queuing in the reverse direction, which may apply on single carriageways where the load forces the road to be closed or obstructs the opposite carriageway. Occasionally, a load may have to run on the ‘wrong’ side for a short time, in response to either manoeuvring or structural constraints. Usually, a section will be blocked completely by police, with a road closure at its downstream end, so the queue will be stationary, which simplifies the calculation. Reverse direction flow can be specified independently or defaulted to be the same as the co-moving demand. This facility is invoked by setting the number of lanes available for passing to a negative value.

3.7 Another facility is to estimate the ‘knock-on’ effect of the discharge flow from the primary queue hitting a bottleneck (assumed fixed) downstream and causing queuing there. The model allows the capacity of the bottleneck to be defined as a proportion of the specified carriageway capacity and also provides for a factor on the discharge flow to allow for dispersion in the intervening distance.

3.8 Finally, a proportion of the queue which has built up on one section can be transferred to the next section. If 100% of a queue is transferred, it means that the sections are fully ‘coupled’. This would be the case if, for example, the change between the sections involved a change in road geometry with no opportunity for traffic to enter or leave, and the load did not pull over the let the queue pass. As shown in Appendices D and E, the queue build-up and delay on two coupled sections is generally much greater than on the two uncoupled sections separately.

4 Calibration and other practical considerations

4.1 The difficult elements in the queue size calculation are the capacity past the load \((\mu_c)\), which cannot be assumed to be the same as for a similar width of road under normal conditions, and the density of queuing traffic \((k_b)\), which has to be obtained as the solution of simultaneous equations involving the speed/flow relationships. There is also a delay from the reduction in the speed of traffic as it passes the load, which depends upon the length of the Zone of Influence, and the speed past the load, both of which may need to be determined empirically. The effect of HGVs should be considered, as they may behave differently from light vehicles in the presence of an abnormal load, and the space for two narrow lanes of cars might be able to accommodate only one lane of HGVs.
4.2 The speed past the load \( v_i \) does not enter into the queuing model because the traffic has already passed the bottleneck at the back of the Zone of Influence. Likewise the channel capacity does not enter into the calculation of passing delay which depend only on speed drop. However, the models assume traffic flow to be smooth and homogeneous, whereas in practice it can be turbulent and heterogeneous. If vehicles in several lanes have to merge to pass the load, they may cause local flow breakdown which reduces their speed and also the effective passing capacity.

4.3 There is anecdotal evidence from the A303, part of which has alternating single and dual sections, that undisciplined or selfish behaviour at merges – overtaking then braking sharply and squeezing in front of a queue - reduces capacity by forcing queuing traffic to slow below its optimum speed. However, this may be less evident when the bottleneck is in motion.

4.4 There is also anecdotal evidence that a high proportion of HGVs could actually increase effective capacity, at least in ‘PCU’\(^3\) terms. Trucks drivers like to maintain speed because it takes a long time for them to accelerate, especially if there is a gradient. Trucks approaching a merge at roadworks on the M2 near Maidstone were observed to line up on two lanes in echelon so as to be able to merge with minimum reduction in speed. This incidentally prevented cars from overtaking, cutting in and causing the flow breakdown mentioned earlier. On the other hand, should HGVs be forced to slow below their optimum speed, especially on an uphill gradient, they may delay overtaking, resulting in an effective lengthening of the Zone of Influence.

4.5 Another cause of slowing could be the HI OCC queue detector\(^4\) which is incorporated in the MIDAS\(^5\) system on the M1, M25 and other motorways. If HI OCC is triggered by a large slow-moving body which registers high occupancy for a long time, even if only in one lane, it will set 40 mph signs which have a dwell time of 4 minutes after the ‘incident’ has passed. If the speed signs are heeded, they could in principle extend the Zone of Influence by about 0.8 mile, in which speeds are slowed to around 40 mph, in addition to the load’s immediate Zone of Influence of say 0.5 mile.

5 Additional contributions to economic cost

5.1 The primary purpose of this report and model is to assess congestion costs. However, the following additional impacts, in approximate order of importance, could in principle be estimated given the variables collected for or estimated by the model, and suitable relationships:

<table>
<thead>
<tr>
<th>Impact</th>
<th>Depends on (X journey time or mileage as appropriate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interference with movement</td>
<td>Delays, diversion, road closures</td>
</tr>
<tr>
<td>Road wear and infrastructure damage</td>
<td>Axle loading, axles, load speed</td>
</tr>
<tr>
<td>Pollution and Climate change</td>
<td>Load speed, weight, delay to other traffic</td>
</tr>
<tr>
<td>Additional accident risk</td>
<td>Queues, difference in speed between load and passing traffic</td>
</tr>
<tr>
<td>Other environmental impact</td>
<td>Road type</td>
</tr>
<tr>
<td>Noise and vibration</td>
<td>Load speed, engine effort related to weight, road standard</td>
</tr>
</tbody>
</table>

5.3 Some published figures for general freight are used to estimate an overall cost contribution of about £3/mile, in addition to comparisons with water and rail (see Appendix I to the main modelling report). However, there is no evidence that these figures are appropriate to Abnormal Loads, nor is there any evidence on the safety impact of queues caused by Abnormal Loads.

\(^3\) PCU stands for Passenger Car Unit, the amount of capacity consumed by a single vehicle relative to that consumed by the average car.

\(^4\) For High OCCupancy detector algorithm – see Collins et al (1979). HI OCC is a fast-acting queue detector which reacts to the occupancy over a single loop detector. It serves as the basis of the queue protection system in MIDAS, setting 40 mph mandatory or advisory speed limit signs.

\(^5\) Motorway Incident Detection and Automatic Signalling system, based on loop detectors.
5.4 Diversion is a possible impact, sensitive to the nature of the surrounding network. Methods of estimating diversion caused by long-term obstructions like road works, such as QUADRO, may not be appropriate to the temporary presence of an AIL, even if warnings are broadcast. Other externalities listed in section 2.4.2 of the Literature Review, such as the cost of providing AIL facilities, are outside the brief of this case-based approach.

5.5 Once conditions are such as to produce queuing, delay rises ever more steeply with increasing demand. Therefore perennial traffic growth can be expected to lead to greater delay for the same movement or to reduced periods in which moves can take place without causing serious disruption. This can be investigated easily by introducing a growth factor into demands, or by drawing demand figures from projected 24h profiles.

5.6 Finally, the economic cost of the delay caused by a load depends on the cost per vehicle-hour, which can vary by day of the week and time of day, according to traffic composition. Up-to-date figures are available from WebTAG (2004). However, the practical variation is small and an average figure can usually be assumed. This is currently £11.28/veh-h for an average vehicle at 2002 prices.

6 Conclusion and results

6.1 This report describes a model for estimating the queuing and delay caused by an abnormal load. The model has been implemented in a spreadsheet which provides for multiple road sections with different characteristics. Three principal types of data are required:

1) Description of the sections making up the route, including road type and number of lanes, normal speed/flow relationships and capacity, and the proportion of flow likely to continue onto next section;

2) Effect of load on each section, including its speed, the speed of passing traffic and number of lanes or capacity available to passing traffic, and the length of the Zone of Influence;

3) Arrival time on each section and normal demand flow and proportion of heavy vehicles (or PCU factor) appropriate to day and time of passage – arrival time may be set automatically to the departure time from the previous section. Diverted volume can also be specified.

6.2 A macro facility aids the process of adding new sections and linking to volume tables taken from TRADS\(^6\) or profiled from ADTs\(^7\), and the selection of data for specific days and start times is largely automated once the model has been constructed, allowing rapid exploration of alternative scenarios. However, obtaining and setting up road section and traffic volume data can be quite labour-intensive and time-consuming.

6.3 The model results are sensitive to certain variables which therefore need careful attention, in particular: normal traffic speeds, volume and passing capacity, especially where demand is near to capacity (it must never exceed normal capacity); speed of the load; start time of journey and time of passage on sections where peak volumes are likely. Another important consideration is how longer route segments are divided up into sections. If queues are unlikely to divert or disperse between these sections, the linkage parameter must be set, but this works correctly only over adjacent section pairs.

6.4 Results are relatively insensitive to the proportion of heavy vehicles, passing speed and the length of the Zone of Influence.

---

\(^6\) Highways Agency's database on [www.trads2.co.uk](http://www.trads2.co.uk), available to registered users.

\(^7\) Average Daily Traffic: a single 24h count available from DfT or various sources
6.5 The results of actually applying the model to various abnormal moves have been encouraging. The example in Appendix H demonstrates the close agreement of model estimates with observed queuing and MIDAS analysis on a section of the M6.

7 Acknowledgements

This work forms part of the Highways Agency project ‘Evaluating Congestion caused by Abnormal Loads’. The author is grateful to original Project Sponsor Nick Wells and Andrew Cook of HA, and Paul Emmerson of TRL, for comments and advice, and to Philip Sanger of TRL for spotting errors in the formulation which have since been corrected, and to Peter Gray for technical advice.

8 References


DMRB Vol 12. Traffic appraisal of roads schemes. DfT/HA
http://www.official-documents.co.uk/document/deps/ha/dmrb/vol12

http://www.official-documents.co.uk/document/deps/ha/dmrb/vol13

http://www.official-documents.co.uk/document/deps/ha/dmrb/vol14


APPENDIX A - TYPES OF QUEUE AND DELAY MODEL

A.1 Dunn’s formula and delay estimation from direct measurements

Dunn (1967) quoted a formula for calculating directly the delay to other road users in vehicle-hours/mile travelled caused by an abnormal load on a single carriageway road – i.e., not allowing any vehicles to pass.

\[
Delay = \frac{n}{V_2} - \frac{1}{V_1}
\]

In this, \(n\) is the mean number of vehicles affected, \(V_j\) is the mean speed (mph) of the vehicles under normal conditions, and \(V_2\) is the reduced mean vehicle speed caused by the presence of the load. In fact, this formula is not helpful for estimation purposes, because it depends on variables which are not straightforward to estimate. An analysis given in Appendix G uses more explicitly measurable quantities, based on actual experience travelling with convoys, to estimate delay provided that queue sizes are small enough to be estimated visually:

\[
\Delta = \frac{\left( L_0 + L_H \right) H}{2v_L} + \frac{N_c Z}{v_c - v_L} \left( 1 - \frac{v_c^*}{v_x} \right)
\]

where \(v_c^* = \max(v_c, v_L)\)

In equation (A2): \(H\) is the section length, \(L_0\) and \(L_H\) are the initial and final queue sizes on the section, \(v_d\) is the prevailing speed of traffic, \(v_L\) is the speed of the load, \(v_c\) is the passing speed, \(Z\) is the length of the Zone of Influence, and \(N_c\) is the number of vehicles which pass on the section (this of course being zero if \(v_c\) is equal to \(v_L\), or it is not possible for traffic to pass). To this should be added any delay to traffic queuing in the opposing direction or on side roads, assuming it can be estimated.

Wherever queues could be substantial, modelling or more extensive observation is likely to be required. The rest of this report is devoted principally to direct modelling methods.

A.2 ‘Vertical’ versus ‘Horizontal’ queue models

Conventional queuing models can be divided into deterministic, where the queue is the accumulated excess of demand over capacity, and steady-state-random, where an equilibrium queue length results from the net effect of random variations in demand and capacity. Time dependent random queuing theory combines these two in a seamless macroscopic model. In dealing with abnormal loads and other obstructions where the important effects are the result of substantial incapacity to serve the demand, especially on high-capacity roads, the simpler deterministic approach is sufficient, though things can get more complex – in a way which is not yet understood – if there are frequent variations in load speed or passing capacity, as can occur on some smaller rural roads.

A more pertinent distinction is between ‘vertical’ and ‘horizontal’ queue models. Vertical models treat the queue as being present at a bottleneck or stop-line, with effectively infinite traffic density in the queue, while horizontal models describe regions of traffic, usually in motion with finite speeds and densities. On motorways and similar roads, queues can extend for great distances and the density and speed of the vehicles in them can vary. Vertical queue models can underestimate both queue size and delay, because they fail to represent the way a growing queue ‘sweeps up’ advancing traffic, and because they do not represent the discharge process accurately, and they also fail to describe the effect on upstream junctions resulting from the physical length of the queue.

The difference between vertical and horizontal queue models is illustrated in Figure 2.
On the left, the vertical model treats the queue as a simple container which fills and then empties. This is equivalent to the head of the queue staying fixed while the tail moves first upstream and then downstream. On the right, the horizontal model recognises that the queue actually discharges from its head while the tail continues to grow, until these processes meet at some point farther upstream.

In the horizontal model the queue discharge time is determined by the physical extent of the queue (eg in km) $X$, which is related to its size (in vehicles) $L$ through the density of traffic in the queue $k_b$ (veh/km), and the speeds with which the boundaries shown in Figure 2 propagate:

$$t_{d(H)} = \frac{X}{v_{ab} - v_{bd}}$$

(horizontal) \hspace{1cm} (T1)

Vertical queue models, on the other hand, derive their discharge time by simply dividing the queue size by the maximum discharge flow, given the prevailing demand which must also be served:

$$t_{d(V)} = \frac{L}{\mu_a - q_a}$$

(vertical) \hspace{1cm} (T2)

During queue growth the delay is the same for both models, but during decay it is proportional to the persistence of the queue, and a horizontal model predicts a longer persistence because of the upstream movement of the tail of the queue. So a vertical model tends to underestimate delay, as well as underestimating how far upstream the queue may be encountered.

In addition to queuing delay there is delay from traffic having to slow down to pass the load. This is the same for both models if it is calculated in relation to the free-flow speed (see later in Appendix B).
A.2 Theoretical models of queuing and capacity

Queue size, in terms of number of vehicles, can be expressed mathematically in the familiar form of ‘demand minus capacity’, with some additional factors which relate respectively to the motion of the load and the finite traffic densities.

The formula (M1) as derived later in Appendix B gives queue size $L$ (assumed to start from zero) as a function of time $t$, in terms of the prevailing speed of traffic $v_{av}$, the speed of the load $v_l$, the channel capacity $\mu_c$ as would be measured by a stationary observer (see Appendix F), and the densities of approaching and queuing traffic $k_a$, $k_b$ respectively, where $v_d$ is the speed of traffic once it has passed the load and has accelerated away downstream (likely to be close to maximum speed):

$$L = \frac{q_a \left( 1 - \frac{v_l}{v_{av}} \right) - \mu_c \left( 1 - \frac{v_L}{v_d} \right) t}{1 - \frac{k_a}{k_b}}$$  \hspace{1cm} (M1)

An alternative model (M2), tested by ITEA, incorporates an adjustment factor for capacity which depends on the ratio of the load speed $v_l$ to the speed of passing traffic $v_c$:

$$L = q_a \left( 1 - \frac{v_l}{v_{av}} \right) - \mu_c \left( 1 - \frac{v_L}{v_c} \right) t$$  \hspace{1cm} (M2)

The capacity adjustment in (M2) is actually that which converts the capacity past the load as measured by a stationary observer to that measured by an observer travelling with the load. The latter is of course zero if $v_c$ is the same as the $v_{av}$, i.e. if no traffic actually passes, but the effective capacity of the road in that case is not zero as (M2) implies, which calls the formula into question.

In both these models, a ‘catch-up’ factor multiplying the demand $q_a$ adjusts it to that which actually reaches the load. For comparison, a conventional vertical queue model can be defined by omitting the density adjustment divisor in (M1) or the capacity adjustment factor in (M2), while retaining the ‘catch-up’ factor to allow for the fact that the load is moving. A further modification which can be applied to all models is to reduce the passing channel capacity, nominally based on the number of free lanes, by a fixed empirical factor such as 0.78, as ascribed to Roberts et al (1994).\(^8\)

To calculate delay it is necessary to take into account in addition to the maximum queue size a number of complex quantities whose full analysis is given in Appendices B-E:

- The total time the queue is present, including while it is discharging;
- Any delay resulting from reduced speed of traffic passing through the Zone of Influence;
- Any additional delay from queues being carried over between adjacent route sections;
- Any additional delay caused by main queue discharge flow hitting a bottleneck downstream.

Horizontal queue modelling requires an understanding of relationships between speed, flow and density, a typical example of which is shown in Figure 4 in Appendix B. There are two distinct and essentially independent regimes, ‘free flow’ in which speed decreases slightly (if at all) as flow

---

\(^8\) $t$ is the period over which the queue develops. Calculation of delay involves a further calculation taking into account the time required by the queue to discharge, and the relative speeds of queuing and free-flowing traffic.

\(^9\) The reference actually asserts that bottleneck capacity at an incident site is reduced on average by 24.2% compared to the free-flow capacity.
increases, as a result of frictional effects; and ‘congested’ or ‘coupled’ where car-following behaviour dominates and flow and speed decrease together. Practical capacity is defined where these meet.

A.3 Comparison of models

Figure 3 compares the results of the models whose variations are described above, for which the table below gives the values of parameters used in the comparison. The model (M2) predicts the onset set of queuing at a lower demand than (M1) and hence initially larger queue sizes. However, once queuing does onset in (M1), queue size rises more rapidly with demand than in (M2) because of the density-dependent divisor. Cost-per-mile is proportional to delay, which depends not only on the size of the queue but also on its persistence, which also differs between the models.

<table>
<thead>
<tr>
<th>Abnormal load parameter</th>
<th>Value assumed for comparison</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lanes in ‘channel’ past load</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Channel capacity past load</td>
<td>4600</td>
<td>PCU/h</td>
</tr>
<tr>
<td>Speed past load (40 mph)</td>
<td>64</td>
<td>km/h</td>
</tr>
<tr>
<td>Speed of load (12 mph)</td>
<td>19</td>
<td>km/h</td>
</tr>
<tr>
<td>Zone of Influence (0.5 mile)</td>
<td>0.8</td>
<td>km</td>
</tr>
<tr>
<td>Length of section (12 miles = 1 hour)</td>
<td>19</td>
<td>km</td>
</tr>
<tr>
<td>PCU factor (15% HGV)</td>
<td>1.225</td>
<td>PCU/veh</td>
</tr>
<tr>
<td>Demand range</td>
<td>200-6500</td>
<td>veh/h</td>
</tr>
</tbody>
</table>

The results show significant differences between the models at high demand levels. The capacity adjustment factor in (M2) has some effect, but it cannot match the steep increase in delay cost in (M1) at high demand levels. It is important to establish the actual capacity past a moving load. There is some support for using the 0.78 factor although it does not strictly apply to abnormal loads. There is also both anecdotal and mathematical evidence that capacity for discharging a slow-moving column of traffic is less than the normal free-flow capacity of high-speed traffic. In addition, driver caution, especially when faced by flashing lights, police practice, merging and ‘rubbernecking’ may account for greater time headways and hence reduced capacity. If the speed of passing traffic falls or is forced below that at which capacity is maximised, then the maximum passing flow will be reduced.

![Figure 3. Comparison of modelled cost-per-mile for a load which leaves 2 lanes free (passing delay, which depends only on passing speed, is 27.7 seconds per vehicle at all demand levels)](image-url)
APPENDIX B – TECHNICAL DESCRIPTION OF THE MODEL

B.1 Speed/flow/density relationships

The flow, speed\(^{10}\) and density of traffic are linked by the so-called ‘fundamental’ relationship:

\[ q = v k \]  

(1)

Various models relating the variables in pairs, such as speed and flow, have been proposed, and most divide the state of traffic into two regimes. At higher speeds a ‘free flow’ regime exists where speed falls moderately, or even stays constant, as flow increases towards capacity. This seems to be determined by the local geometry and some ‘friction between vehicles’. In the ‘congested’ regime, perhaps better described as ‘coupled’ by car-following behaviour, flow falls as speed falls. This normally arises where there is a capacity restriction downstream\(^{11}\). Observations confirm that these regimes are qualitatively different. A simple model was suggested by Duncan (1976, 1979) who fitted an almost linear free-flow relationship, and a coupled model of fixed minimum distance headway plus fixed time headway, as in Figure 4\(^{12}\). A similar conclusion was reached by Banks (1989). Standard COBA speed/flow relationships apply to the free-flow regime, but the ‘channel’ left by an abnormal load may have special properties which need to be calibrated from experience. COBA relationships often have a bi-linear form, rather than the linear form of the upper graph in Figure 4, and other models exhibit curvature of the free-flow relationship.

\[ v = 120 - \frac{x}{60} \]

\[ x = \frac{3000}{1 + \frac{y}{30}} \]

(Headway = 6.33m + 1 sec)

Curve fitted by eye

Figure 4. Speed/flow relationships fitted directly to divided speed/flow data (Duncan 1979)

\(^{10}\) Speed in this context is space-mean rather than time-mean, though the distinction is less important when dealing with aggregate quantities and long time periods. Space mean speed \(v_s\) is defined in terms of time-mean speed \(v\) and its standard deviation \(\sigma\) by the formula: \(v_s \approx v / \left(1 + \sigma^2 / v^2 \right)\)

\(^{11}\) Referring to a ‘congested’ regime can be misleading because traffic can flow quite fast, although it may become unstable

\(^{12}\) The maximum lane flow achieved in the Figure seems surprisingly high, but it can be taken as illustrative.
Speed/flow states can arise from a number of causes, and trying to ‘fit’ a model to heterogenous data may reduce its explanatory value. Non-linearity in the free–flow relationship could be a reflection of localised car-following bringing a coupled relationship into play, as at the top end of the lower graph in Figure 4, or could be associated with traffic discharging from a queue, which tends to maintain a constant flow near capacity over a range of speeds (Navin and Hall 1989). The coupled relationship in the lower half of Figure 4 will be assumed to depend only on car following behaviour in a single stream of traffic, as proposed by Duncan (1979) and Banks (1989):

\[
q = \frac{v}{\tau' v + \lambda'} = \frac{1 - \lambda' k}{\tau'}, \quad v = \frac{\lambda' q}{1 - \tau' q}, \quad k = \frac{1}{\tau' v + \lambda'}
\]

where \(\tau' = \frac{\tau}{n}\), \(\lambda' = \frac{\lambda}{n}\), \(n = \text{number of lanes}\)

The parameters \(\tau\) and \(\lambda\) are nominal time and distance headways, which in reality probably depend somewhat on speed and road geometry, and possibly on traffic composition, but are treated here as constant. The point where the free flow relationship meets the coupled relationship determines the effective road capacity, ignoring any ‘overhang’ in the free-flow regime which may reflect lower capacity for coupled traffic associated with flow breakdown. This point is very sensitive to the parameters \(\tau, \lambda\), so it is best to calibrate them to the known capacity, say \(\mu\).

The simplest model of free flow is a linear speed/density relationship such as Greenshields’ (1935):

\[
\frac{k}{k_m} = 1 - \frac{v}{v_m}
\]

where the ‘jam density’ is \(k_m\) and the free speed is \(v_m\). This can also be applied to the ‘coupled’ regime, and together with (1) leads to a parabolic shape of speed/flow relationship. Figure 4 and similar data do not support this shape, although it has been suggested that data might be fitted using parts of two parabolic relationships (Duncan). Under free flow only, since the speed/flow relationship tends to be quite linear as well as fairly flat in most cases, for ease of calibration and to get explicit solutions it will be convenient to assume here the following linear relationship:

\[
v = v_m - \alpha q
\]

The parameters \(\alpha\), \(\lambda\), and \(\tau\) are difficult to measure directly, but it is relatively easy to obtain average values for \(v_m\) and maximum flow \(q_m\) or \(\mu\), by direct observation. As will be shown later, the ratio \(\lambda/\tau\) is in theory the speed at which a shock wave will travel upstream causing traffic to stop momentarily. This again can in principle be measured directly for any road, and on motorways is consistently around 19-20 km/h. From these parameters it is possible to calibrate the formulae to give an approximation to observed speed/flow relationships.

B.2 Speed and flow in the vicinity of a load

Figure 5 shows the situation of a moving load with a passing ‘channel’, as seen by two stationary observers.

---

13 Although such average values may not be reliable or consistent.
Figure 5. Observation of traffic flow near load over a short period

O₁ is ahead of the load\(^{14}\) throughout the period \(t\), and therefore counts throughput \(q_{c,t}\), where \(q_{c}\) is the flow past the load, which may or may not equal the effective capacity. Note that this flow is the static capacity and does not depend explicitly on the speed of the load, although this could affect its practical capacity through altering the behaviour of drivers. O₂ is behind the load throughout the period, and therefore counts the net change vehicles which occupy the space which is behind the load at the end. O₂ also measures the actual throughput of the upstream traffic, \(q_{u,t}\). Provided that conditions are stable, these measures can be taken to apply throughout the load’s journey along a road section. Alternatively, several sets of observations over short periods can be combined to establish a relationship between average measurements. The flow rates therefore satisfy:

\[
q_{b} = q_{c} + (k_{b} - k_{c})v_{L}
\]  

(5)

This is a form of a familiar equation which reflects conservation of vehicles, the load acting as an extended boundary between the queuing and downstream regions. Using (1) and rearranging:

\[
q_{b} = \frac{q_{c}F}{1 - \frac{v_{L}}{v_{b}}} \quad \text{where} \quad F(q_{c}) = 1 - \frac{v_{L}}{v_{m} - \alpha q_{c}}
\]  

(6)

Note that the speed in the denominator of the expression for \(F\) is the speed of free-flowing traffic downstream, which is likely to be approaching the maximum speed on the road, not the speed at which traffic passes the load, which is likely to be much lower. In the case of a stationary bottleneck, \(v_{L} = 0\), we get \(q_{b}=q_{c}\), ie the flow upstream of the load is equal to the capacity of the channel, as expected. With a moving load, the normal speed/flow relationships governing the upstream traffic apply. If no traffic can pass the load, so that \(q_{c}=0\), then \(v_{b}=v_{L}\) and (6) gives no information.

The factor \(F\) has only one value of practical interest, that where passing flow \(q_{c}\) equals passing capacity \(\mu_{c}\), so it can be treated as a constant adjustment. Generally, if the traffic is free-flowing then \(q_{b}=q_{a}\) and \(v_{b}=v_{a}\), which depend only on the given demand, so the actual flow past the load is determined explicitly. Thus traffic will be free-flowing provided that this flow, sometimes referred to as ‘adjusted demand’, does not exceed the channel capacity modified by the factor \(F\):

\[
q_{a} \left(1 - \frac{v_{L}}{v_{a}}\right) < \mu_{c}^{*} \quad \text{where we define} \quad \mu_{c}^{*} = \mu_{c}F(\mu_{c})
\]  

(7)

Otherwise, an explicit value for \((v_{b},q_{b})\) can be obtained by solving (6) with (2), assuming \(q_{c}=\mu_{c}\). Equation (6) has a hyperbolic shape with negative slope, the opposite of a normal ‘coupled’ speed/flow relationship, having minimum speed \(v_{L}\) and minimum flow \(q_{c}\), as shown in Figure 6. Possible traffic states upstream of the load are the points where this contour meets the prevailing

\(^{14}\) Or more precisely ahead of the back end of the load’s zone of influence, where the effective bottleneck is.
speed/flow relationships. It is interesting that in the case of a moving load the free-flow and coupled solutions occur at different flow values.

The solution is:

\[ q_b = \frac{v_L + \lambda^* \mu^*_c}{\mu^*_c} \quad \quad v_b = \frac{v_L + \lambda^* \mu^*_c}{1 - \mu^*_c} \quad \quad k_b = \frac{1 - \mu^*_c}{\mu^*_c v_L + \lambda^*} \]  

(8)

The flow in the queue, \( q_a \), is the effective capacity of the moving bottleneck. The effect of load speed on this value can be spectacular, as shown by Figure 7.

One way to obtain a measure of the queue size at any time is to track the boundary which forms between the demand flow and the traffic upstream of the load, which is effectively the back of the queue. Since this is unconstrained, it moves at the speed given by equation (9) below, a common formula which arises from conservation of traffic at a phase boundary and does not depend on any speed/flow relationship:

\[ v_{ab} = \frac{q_a - q_b}{k_a - k_b} \leq v_L \]  

(9)

Note that the speed of this boundary cannot exceed \( v_L \), otherwise it would catch up with the load and the queue would disappear. It must be remembered that evaluations of (8) and (9) are meaningless unless a queue is present as determined by equation (7).
Figure 7. Effect of load speed on effective bottleneck ‘capacity’, assuming a normal carriageway capacity of 7200 veh/h, channel capacity of 1800 veh/h, and the Duncan/Banks coupled flow model

B.3 Queue caused by the load

When calculating the queuing delay, it is necessary to consider both the build up and discharge of the queue. This is idealised in Figure 8.

Figure 8. Horizontal queuing model based on boundary wave speeds
From the geometry of Figure 8:

\[ X = (v_L - v_{ab})t \quad (10) \]

\[ L = Xk_b \quad (11) \]

\[ t_d = \frac{X}{v_{ab} - v_{bd}} \quad \text{and total duration of queue} \quad t_c \equiv t + t_d \quad (12) \]

Unlike total delay, which is unambiguous, queue length depends on when, where and how it is measured. This can become complicated where there is blocking back of the queue between road sections. An *ad hoc* method of evaluating representative queue lengths is described later in Appendix E, along with a queuing model extended to deal with blocking back over many sections.

### B.4 Interpretation of queue in conventional demand and capacity terms

Taking equations (8-11) together, with some rearrangement and substitution to eliminate \( q_b \) in favour of \( \mu_c^* \), the adjusted static channel capacity as defined by (6) and (7), the following expression for queue size is obtained:

\[ L = \left( \frac{q_a \left( 1 - \frac{v_L}{v_a} \right) - \mu_c^*}{1 - \frac{k_a}{k_b}} \right) t \quad (13 \text{ or M1}) \]

The numerator has the conventional deterministic form of demand minus capacity, while the divisor represents a ‘horizontal’ correction depending on the relative densities of the arriving and queuing traffic. The effect of the moving load is accounted for by adjustment of the demand and the channel capacity. By rearranging the formula, it is possible to absorb the capacity adjustment factor into other terms, so it is still possible to talk about the ‘capacity’ of the system being the physical capacity of the channel. If \( v_L \) is zero then demand reverts to simply \( q_a \) and capacity to \( \mu_c \), and if the model is ‘vertical’ then the denominator reduces to unity, restoring the conventional deterministic queue model.

### B.5 Total delay caused by the load

Unlike in the vertical case, the delay caused by the queue depends on the difference in the speed between free-flowing and queuing traffic. A column of traffic moving at normal traffic speed incurs no delay, but queuing traffic which is moving incurs less delay that stationary traffic. From the geometry of Figure 8, total queuing delay is given by the equation (14):

\[ D_b = .5Lt_c \left( 1 - \frac{v_b}{v_a} \right) \quad (14) \]

However, this equation takes no account of blocking back, or linking of the queue between sections, so in practice requires extensive modification as detailed later in Appendices D and E, the first of which describes *ad hoc* adjustments which are no longer used but are included for completeness.

Delay is also incurred in passing the load, but the vehicles affected by queuing and passing delay on each section are not in general identical. Ignoring deceleration and acceleration, or absorbing their
contribution into the effective ‘Zone of Influence’, length $Z$, of the load, the time taken by one vehicle to pass the load, assumed to be subject to a maximum of the time spent by the load on the section, is:

$$ t_c = \min \left( \frac{Z}{v_c - v_L} , t \right) $$ (15)

and the net delay ‘in the channel’ is the excess of this over the time taken to travel the same distance at the normal free-flow speed $v_a$ corresponding to the demand flow $q_d$:

$$ d_c = t_c \left( 1 - \frac{v_c}{v_a} \right) $$ (16)

The total delay resulting is imposed only on the traffic which actually passes the load:

$$ D_c = d_c q_c t $$ (17)

The total delay to which traffic is subjected is then:

$$ D = D_b + D_c $$ (18)

Another way of looking at delay is in terms of cumulative flow downstream of the load, as shown in Figure 9. In principle, total delay can be measured directly, by comparing the flow profile downstream of the load with the flow profile which would be expected under normal conditions, provided that this continues long enough for any queue to disperse and flow to return to normal, and that no traffic ‘escapes’ from the section.

![Figure 9. Relationship between total delay and cumulative downstream flow](image-url)
B.6 Estimating average delay per vehicle

The average delay per vehicle affected by the load can be an important result as it reflects the effect of the load as perceived by an individual driver, and therefore may enter into considerations about whether the delay would be perceived as acceptable or unacceptable. While passing delay is already calculated per vehicle by equation (16), and can affect each vehicle only once, calculating queuing delay per vehicle requires values for the number of vehicles, bearing in mind that the same vehicle could be affected on several road sections. It is straightforward to calculate number of vehicles affected for an isolated simple queue, as follows, where the factor involving $v_{ab}$ accounts for the tail of the queue ‘sweeping up’ extra vehicles:

$$Q_b = q_a f_e \left(1 - \frac{v_{ab}}{v_a}\right) \quad (19)$$

$$d_b = \frac{D_b}{Q_b} = \frac{5L}{q_a} \left(1 - \frac{v_b}{v_a}\right)/\left(1 - \frac{v_{ab}}{v_a}\right) \quad (20)$$

Calculation becomes problematic when queues spill back over several road sections, because the same vehicle may be affected by the queue on more than one section, and simply adding up (19) for all the sections results in multiple counting. Some attempts to estimate the number of vehicles affected on a section basis have produced counterintuitive results. For this reason, further discussion is postponed until Appendix E, after several other essential elements and concepts have been described.

B.7 Linking of queues between successive route sections

‘Linking’ here means that all or part of a queue formed on one section follows the load onto the next section. This is the converse of ‘blocking back’ or ‘spill-back’. Analysis for two linked sections can be difficult, and for more than two becomes extremely complicated. In the present model, a practical requirement is to be able to evaluate a route section by section, with minimal dependence between adjacent sections. Several approaches have been investigated, culminating in an extension to the basic queue model. **These are detailed later in Appendices D and E.**

In all cases, any traffic following the load onto a new section modifies the demand on that section as shown in Figure 10, where the proportion of traffic on section 1 which continues onto section 2 is $p$. The model is not designed to cope with pre-existing congestion, so the normal demand on section 2, $q_{a2}$, should be within the capacity of the section. However, this need not be the case for the total demand when a queue is present upstream. Excess demand will cause a queue to develop on section 2, which may spill back onto section 1 and any entering arms. Although traffic on entering arms is not modelled explicitly, it can be taken into account by considering the flow balance at the intersection and how this is affected by turning traffic and queuing.
This demand, given by equation (23), depends on whether the component from section 1 is in free flow, \( q_d \), or is queuing, \( q_b \):

\[
q_{a2(l)} = q_{a2} + p(q_{d1/b1} - q_{a1})
\]  

(21)

One way to determine which section 1 flow to use is to evaluate (19) for free flow and calculate the corresponding tail wave speed, using the standard formula (as in equation (9)). If this would overtake (ie is more negative than) the discharge wave on section 1, then the hypothesis of free flow must be false, so the queuing flow on section 1 should be used, the corresponding ‘tail wave’ being now the change wave between the traffic queuing under the influence of the load on section 1 and that queuing under the influence of conditions on section 2.

B.8 Knock-on effect of released queues

The discharge flow from a queue is assumed to occur at capacity rate. It has been observed that this flow can be maintained for a considerable distance downstream, with little dispersion. If such a heavy flow hits a bottleneck where capacity is reduced then it may trigger flow breakdown and queuing, which would not have occurred otherwise, and therefore can be ascribed to the load. The preceding period of reduced flow in the ‘shadow’ of the load will not compensate for this, although it could delay the onset of any normal queuing. The delay caused by this knock on effect can be estimated most simply by projecting the main queue discharge flow, with optional dispersion, onto a fixed bottleneck downstream of specified capacity (most conveniently expressed as a fraction of the normal carriageway capacity), as illustrated in Figure 11. The profile of the queue resembles Figure 3 (left) because it is the demand which eventually falls, not the capacity which is restored. Some dispersion of the flow may be expected, and this is taken into account by a factor. For the sake of simplicity any traffic which joins or leaves the carriageway in the interim is ignored. The bottleneck parameters will require calibration from actual data related to specific locations and circumstances.

Total delay in this case can be calculated by an equation similar to (14), subject to speed correction, but the time during which the queue is growing is less than the (dispersed) duration of the increased demand, depending on the relationship between the flows, densities and speeds of the arriving and queuing traffic. The former are obtainable as a by-product of the main queue and delay calculation, while the latter are obtained from a speed/flow relationship consistent with the bottleneck capacity \( \mu_Q \). The standard carriageway speed/flow relationship is assumed to apply in the body and upstream of the queue, from which the speed \( v_Q \) and density \( k_Q \) can be derived.
The speed $v_{Q_d}$ of the back of the knock-on queue (negative meaning upstream-moving) is given by:

$$v_{Q_d} = \frac{\mu_Q - q_d/f}{k_Q - k_{d(f)}}$$

(22)

where subscript $d$ refers to the arriving traffic, $f$ is the dispersion factor, and $(f)$ in a subscript means the quantity applies after dispersion, though no correction may in fact be necessary. The actual queue growth time is now given by:

$$t_{Q_d} = \frac{t_d v_d f}{v_{d(f)} - v_{Q_d}}$$

(23)

The maximum queue size in vehicles is now given by:

$$L_Q = -k_Q v_{Q_d} t_{Q_d}$$

(24)

Discharge of this queue is assumed to depend in a simple way on the spare capacity after satisfying the prevailing demand, which is assumed to be less than the bottleneck capacity, ie queuing would not normally occur:

$$t_{Q_d} = \frac{L_Q}{\mu_Q - q_d}$$

(25)

Finally the total additional delay from knock-on queuing is calculated, taking into account the difference in speed from the arriving traffic:

$$D_Q = 5L_Q \left( v_{Q_d} + t_{Q_d} \left( 1 - \frac{v_Q}{v_{d(f)}} \right) \right)$$

(26)

### B.9 Modelling the effect of diversion

The QUADRO software (see DfT websites) applies the well-established equilibrium principle that traffic will divert until the travel time along the diversion route equals that along the original route. This requires detailed knowledge of the diversion route(s) and how travel time on them is likely to be
affected by increased traffic. It also assumes that drivers have perfect knowledge of any delays, which is feasible for the protracted roadworks cases for which QUADRO was developed, but may not be appropriate for ‘one-off’ abnormal load moves.

Analysing possible diversion routes is complex but in many cases diversion can be assumed to be small – for example, diversion of heavy motorway flows is generally impractical and is not encouraged by the police, and routes through hilly rural areas tend not to offer convenient diversions. In other cases, the model allows a simplified application of the equilibrium principle. By entering according to judgment a diverted volume on each affected section, which must be less than the normal volume, the actual volume following the load is reduced, and the queues and delays reduced accordingly. Using the delay and cost per vehicle so obtained, the model then automatically factors up the total delay and cost according to the ratio of total to undiverted flow, on the assumption that all vehicles experience the same delay or equivalent opportunity cost.

B.10 Summary and calibration of model

The model described has isolated the components of delay caused by an abnormal load, expressing them in terms of some basic traffic flow parameters. This model is a horizontal queue model which takes account of traffic speeds and densities as opposed to a vertical model which treats queues as if they are concentrated at the bottleneck.

Three principal types of data are required in the spreadsheet implementation:

1) Description of the sections making up the route, including road type and number of lanes, normal speed/flow relationships and capacity, and the proportion of flow likely to continue onto next section;

2) Effect of load on each section, including its speed, the speed of passing traffic and number of lanes or capacity available to passing traffic, and the length of the Zone of Influence;

3) Arrival time on each section and normal demand flow and proportion of heavy vehicles (or PCU factor) appropriate to day and time of passage – arrival time may be set automatically to the departure time from the previous section. Diverted volume can also be specified, as described under B.8.

The model results are sensitive to certain variables which therefore need careful attention, in particular: normal traffic speeds, volume and passing capacity, especially where demand is near to capacity (it must never exceed normal capacity); speed of the load; start time of journey and time of passage on sections where peak volumes are likely. Another important consideration is how longer route segments are divided up into sections. If queues are unlikely to divert or disperse between these sections, the linkage parameter must be set, but this works correctly only over adjacent section pairs. These are the variables most likely to need calibration when aiming to match or explain observed data. Results are relatively insensitive to the proportion of heavy vehicles, passing speed and the length of the Zone of Influence.

Results are fairly insensitive to the speed/flow parameters assumed for each section, provided that (i) capacity is sufficient, (ii) free speed and speed at capacity are realistic, and (iii) the speeds of the load and passing traffic are compatible with the normal speeds on the section. The COBA Manual, DMRB Vol 13, contains detailed speed/flow functions regressed on various geometric and visibility parameters for various types of road under normal flow condition. In the ‘coupled’ or car-following regime which applies to traffic queuing behind the load, provided that the correct normal carriageway capacity and speed-at-capacity are calibrated, suitable values of the parameters \( \tau \) and \( \lambda \) can be derived automatically. Speed/flow relationship applying to passing traffic are not known, so speeds and especially capacities need to be based on observed data, and the car-following behaviour of passing traffic, which could be different from that of normal traffic, is implicit in these values. Part of the Project is to obtain the necessary observational data.
APPENDIX C – MODELLING THE EFFECT OF BREAKS

Drivers of Abnormal Loads have to take 45 minutes of breaks in every 4½ hours. Practically, such breaks need to be at least 15 minutes long, but hauliers also desire to make steady progress so breaks are unlikely to be much longer than this unless actual repairs are needed. Since loads cannot stop ‘in the middle of the road’, breaks must be taken at places where it is known that the vehicle can be accommodated without causing an obstruction to passing traffic.

C.1 Modelling breaks

Modelling of breaks is in a sense the converse of modelling queue linking but is simpler because when dividing a single section it is reasonable to assume that the parts have similar characteristics. It is not necessarily correct, but is reasonable for the sake of simplicity to assume also that the parts are equal. It is straightforward to calculate the effect of $n$ evenly-spaced breaks each of which allows the queue to discharge completely, since this reduces the maximum queue and total delay by the factor of $n$. However, it is necessary to take account of the actual length of breaks, since if they are not long enough to allow the queue to pass completely, then delay will be only partially reduced.

Another factor is what proportion $p$ of the queue is assumed to be passed on to the next section (see earlier in Section B.6). A break at the end of a road section will have no effect on delay if no vehicles would have followed the load onto the next section anyway. It is reasonable to assume that stops are positioned so that they provide some benefit to other traffic. For example, if no queue is carried over, then the optimal placing of two breaks might be at 1/3 and 2/3 along the section, while if all the queue is carried over, then the optimal placing might be at 1/2 way along and the end of the section. This can be represented by placing the breaks at:

$$b_i = \frac{i}{n+1-p} \quad (i = 1 .. n) \quad \text{(C.1)}$$

From the viewpoint of queue calculation, the section is then divided into $m$ parts, where:

$$m = \max\left[ (n+1-p)\left(1+\frac{n}{2}\right) \right] \quad \text{(C.2)}$$

The time needed for the queue to pass on the undivided section is greater than the discharge time because the head of the queue moves upstream and extra time is needed for the dense discharge traffic to pass the bottleneck. This extra time is assumed to be that required to cover the distance involved at the ‘capacity speed’ of the carriageway.

$$t_p = t_d + \left(\frac{Z - v_{bd}t_d}{v_q} \right) \quad \text{(C.3)}$$

If the total stop time on the section is $T_S$ (including a stop specifically at the end of the section) then the proportion of the present queue which is able to pass at each break (or the end of the section) is:

$$b = \min\left[ \frac{mT_S}{t_p \max(n,1)}, 1 \right] \quad \text{(C.4)}$$

Note that according to this definition, $m$ need not be integral.
Since this proportion applies to each of the breaks, the factor by which maximum queue length and total delay are adjusted, and the corresponding delay saving on the kth road section are:

\[ f = \frac{1}{1 + (m - 1)b} \]

\[ \Delta_\chi = (1 - f)D_{bk} \quad (C.5) \]

The effective proportion of unmodified queue carried over to the next section is given by:

\[ p^* = p(l - b) \quad (C.6) \]

This model is quite simple to implement and easier to set up than an equivalent model linked sections, and is valid for any number of breaks in a section. Furthermore, using a similar approach and given a route and its timings, it is possible to estimate where and how many breaks of a specified length (eg 15 minutes) would be most beneficial.

C.2 Practical recommendation

The model described calculates a number of breaks for each road section, but it takes no account of either the relative value of these breaks, which can vary widely, or the practical number of breaks which can be taken on a journey, for which a target value might be based on the ‘45 minutes in 4½ hours’ rule. On the other hand, creating an optimum schedule of breaks would involve a level of judgment arguably beyond the remit of the model, as well as a complexity of calculation which may not be possible in a spreadsheet program.

As a compromise, it is noted that justification for selecting a subset of stops is most likely when there is a large number of potential stops and great variation in their values. The following formula calculates the number of breaks which would be justified by the potential delay saving on each section, relative to the total number of breaks and delay saving estimated by the primary model:

\[ n_{k_\text{rec}} = \min(n_k, n_{k-}) \quad (C.8) \]

The ‘target’ number of breaks is obtained in proportion to the total travel time according to the rule given earlier, with rounding (so, for example 3 15-minutes breaks would be targeted for travel time in the range 4-5 hours). If the total number of potential breaks \( \Sigma n_k \) exceeds this, then the recommended number of breaks for each section is calculated as below, and the delay saving estimated pro rata:

\[ \Sigma n_{k_\text{rec}} \leq \Sigma n_k \]

Since \( \Sigma n_{k_\text{rec}} = \Sigma n_k \), if there are sections with high value breaks, for which \( n_{k_\text{rec}} \gg n_k \), then there are likely to be other sections for which \( n_{k_\text{rec}} = 0 \). By (C.8), no breaks will be recommended for these, so \( \Sigma n_{k_\text{rec}} < \Sigma n_k \) and higher value breaks will be favoured. Practical tests have been run with three models, with the following results which appear to confirm expectations:

<table>
<thead>
<tr>
<th>Table C.1. Potential and recommended breaks for some moves</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of move</strong></td>
</tr>
<tr>
<td>Girder-frame via Mway</td>
</tr>
<tr>
<td>Move via mixed D2-D4</td>
</tr>
<tr>
<td>Move mostly via Mway</td>
</tr>
</tbody>
</table>
APPENDIX D – INTERPOLATIVE MODELS OF LINKED SECTION QUEUES

This Appendix reports early models of delay involving queues linked between road sections (i.e. blocking back) for the sake of completeness. The objective was to set up a recursive procedure simple to implement (even if the formulae could be quite ‘heavy’). These models have been superseded by the model described in Appendix E.

D.1 First model of linking between pairs of sections

As explained earlier in Section B.6, a practical requirement is to calculate delay along a load’s route section by section, and this is really precludes exact calculation of queuing linked over several sections. Several approaches have been developed, and are described here for the sake of redcord, and also to illustrate the pros and cons of simplified approaches. The first model developed is described here. It deals with pairs of consecutive sections only and uses a general approach to estimate the extra delay on a section by apportioning the total extra delay which would occur if the sections were fully coupled. This was acceptable at the start of the project because of the long sections then used in modelling. It must be emphasised that it is a much simplified method whose primary purpose is to give the correct answer when a single homogenous section is split in two at an arbitrary point and no vehicles are able to enter or leave at that point. Three cases are distinguished, as in Figure D.1:

![Diagram of three cases of coupled queues on two sections, and definition of flow proportions](image)

The delay produced by a queue is proportional to the area it occupies on a space-time diagram, or more precisely on a vehicle-time diagram, the two being connected by the spatial density of vehicles. In this case, for the sake of simplicity, we have used distance as the vertical axis. For a coupled queue, the delay is proportional to the area of the large triangle, while for uncoupled queues the respective delays on the two sections are proportional to the areas of the white triangles. The potential extra delay associated with coupling is therefore proportional to the shaded area. It is potential because it...
applies only if all vehicles queuing on section 1 move onto section 2. If this is not the case then the actual end of the queue on section 2 will not coincide with the head of the queue on section 1. However, we assume that the actual delay in all cases can be obtained from the potential delay by simply multiplying by the flow factors $p_1$ and $p_2$ defined in the lower right part of Figure D.1, and adjusting for the difference in speed between queuing and normal traffic (see earlier in Section B.3).

Note that all the triangles are similar, a fact which we can use to simplify the estimation. If the total potential delays on the independent sections are $D_1$ and $D_2$ respectively, then the potential total and extra delay for coupled sections are given by:

$$D_+ = \left(\sqrt{D_1} + \sqrt{D_2}\right)^2$$
$$\Delta_+ = 2\sqrt{D_1D_2}$$  \hspace{1cm} (D.1)

In the first two cases (top of Figure D.1), the back of the queue moves downstream, and the extra delay can be apportioned between the two sections on the basis that we wish to estimate delay on each section regardless of where the cause lies. In the third case (lower left) the back of the queue moves upstream and all the extra delay occurs on section 1. In the first case (top left), the potential extra delay on section 1 is easy to calculate from geometry as (see Section B.3 and Figure 8 for definitions):

$$\Delta_1^* = .5H_1^2k_{pl}\left(\frac{1}{v_{abl}} - \frac{1}{v_{L1}}\right) - D_1$$  \hspace{1cm} (D.2)

In the second case this exceeds the potential delay on section 1, but the whole of the triangle on section 2 is potential delay, and the potential extra delay is given by:

$$\Delta_2^* = .5H_2^2k_{b2}\left(\frac{1}{v_{L2}} - \frac{1}{v_{bd2}}\right) - D_2$$  \hspace{1cm} (D.3)

From the Figure geometry, the extra delay on section 1 given by (D.2) will be the true value provided that no more than half the total extra delay occurs on section 1. Otherwise, the extra delay on section 2 will be the true value and must be subtracted from the total to give the section 1 delay. In summary:

(a) If $v_{abl} \leq 0$:  
$$\Delta_1 = \Delta_+$$  
$$\Delta_2 = 0$$

(b) otherwise, if $\Delta_1 \leq .5 \Delta_+$:  
$$\Delta_1 = \Delta_1^*$$  
$$\Delta_2 = \Delta_+ - \Delta_1^*$$  \hspace{1cm} (D.4)

(c) otherwise:  
$$\Delta_1 = \Delta_+ - \Delta_2^*$$  
$$\Delta_2 = \Delta_2^*$$

Finally, we can estimate the actual extra vehicle-hours delay on the sections as follows:

$$\Delta D_{bl} = p_1\left(1 - \frac{v_{bl}}{v_{a1}}\right)\Delta_1$$  
$$\Delta D_{b2} = p_2\left(1 - \frac{v_{b2}}{v_{a2}}\right)\Delta_2$$  \hspace{1cm} (D.5)

Estimating the effect on maximum queue length *ad hoc*, treating this as independent of delay, in the upper two cases in Figure D.1 only the maximum queue length on section 2 is affected, and in the lower case only that on section 1. The effect can be estimated mostly from geometry and is most easily done in terms of physical extension. If the back of the queue on section 2 is moving upstream
then the uncoupled queue length there could be calculated to be greater than the length of the section, in which case the original queue length can be retained. Thus:

If \( X_2 \leq H_2 \):

\[
X_2^\Delta = \min \left[ (1 - p_2)X_2 + \frac{p_2X_2 \sqrt{D_2}}{\sqrt{D_1}}, H_2 \right]
\]  
(D.6)

For the third case, the extended queue on section 1 is trickier to estimate since it could in principle reach a maximum extent when in a different flow condition to when it was directly behind the load. A possible correction is:

If \( v_{\text{abl}} \leq 0 \):

\[
X_1^\Delta = \max \left[ p_1 \left( X_1 - v_{\text{abl}}H_2 \left( \frac{1}{v_{L2}} - \frac{1}{v_{bd2}} \right) \right), X_1 \right]
\]  
(D.7)

The advantage of this method compared to earlier methods tried is that it does not make use of the queue sizes calculated for uncoupled sections, making it possible to derive equations (D.6) and (D.7) without circularity. However, other methods are possible\(^{16}\), and it is not known how well this method deals with cases where the section characteristics are substantially different, and/or the performance of the load on them is substantially different.

### D.2 Problems with the two-section method

The method described above has turned out to be unsatisfactory since the overall model was first implemented, as it has become common to break down motorway runs into their individual junction-to-junction sections each of which has its own TRADS volume table or ADT value. This means that in most spreadsheet models involving a motorway run, a large number of consecutive motorway sections may occur, and if traffic is heavy queues could be linked over many more than two sections.

A feature of the two-section model is that it attempts to apportion delay and queue length, expressed either as extent in kilometres or size in vehicles, between the two sections. However, if a queue on the upstream or an isolated section would spill back this is not accounted for, so in that sense the model is not complete. A further point is that where delay is incurred may be less important than where it is determined – i.e. which road section would, if removed, also remove the delay. Accounting delay in this manner avoids the virtually intractable\(^{17}\) problem of apportioning between several sections. Furthermore, the model is exact only for similar sections, where the road, load and volume characteristics are the same except for section length, which is unlikely to be the case in practice. Finally, it is unlikely that precise factors for the proportion of queue passed on from one section to the next can be obtained.

Extension of the two-section model to multiple sections is possible, provided that apportioning of delay is dropped. However, this still requires a matrix or iterative calculation. A more promising approach is to recognise that in a linked or blocking-back situation, the arrival rate on each section is not constant but involves at least two different flow rates, one of which includes outflow from the queue upstream. In the space/time diagram, this results in a quadrilateral instead of triangular queuing

---

\(^{16}\) The method described here is referred to (arbitrarily) as ‘Model Q’ in some spreadsheet calculations. The method referred to as ‘Model P’ was based on queue sizes and factors derived from boundary speeds. For sections with identical characteristics, it produces the same extra delay estimates as the present model, although the latter tends to ascribe more delay to the first section where technically it occurs.

\(^{17}\) In practice requiring matrix calculations or iteration, and not calculable on a section-by-section basis, even with cross-referencing of variables between adjacent sections. This occurs because the result on a given section may depend on values from all the sections to which it is linked, so for \( N \) sections \( N(N+1)/2 \) calculations are required, whereas a single- or two-section model requires \( N \) or \( 2N \) only.
region. It turns out that this kind of model can calculate delay in a way which appears realistic on a section by section basis, giving the known correct results for similar sections.

As regards queue length, whereas delay is a ‘commodity’, in the sense of a physical, additive quantity whose magnitude does not depend on how it is measured, queue length is meaningful only as measured at a particular moment (extended over space) or a particular place (extended over time), and queue lengths on different sections of a route are not additive because the sections are traversed at different times. Therefore, more notional estimates of queue length should be acceptable, again avoiding a problem of intractability.

**D.3 Extension to more than two sections**

In Figure D.2 below, which is a highly schematic diagram of a multi-section queue, the queuing regions calculated by the two-section model are shown unshaded. The triangular regions with bold edges are the queues calculated by the basic single-section model, and would be the queues which would occur if there were no linking. Figure D.2 represents a special and artificial case in which the tail of each queue remains stationary, avoiding the issue of apportioning. Where more than two sections’ queues are linked, as shown, the total delay is underestimated because the regions labelled 13, 14 and 24 (heavily shaded) are left out. This omission gets worse as the number of linked sections increases. With this simple configuration, true delay is broadly proportional to the square of the number of sections, whereas with the two-section model, although initially correct for two sections, it tends towards a value proportional only to the number of sections.

![Figure D.2 Queuing regions covered by single- and two-section models](image)

A multi-section model can be defined by extending the model described in Section D.1. For reasons given earlier, this is not considered a practical solution for spreadsheet implementation, but is instructive and included for the sake of completeness. In accordance with the reasoning in section D.1, if the potential delays on two similar consecutive road sections $S_1$ and $S_2$ are $D_1$ and $D_2$ respectively, then the potential total and extra delay for the sections when all queuing traffic on the
first passes onto the second are given by the following, where the square-root-of-delay terms act as a proxy for the role of queue size:

\[
D_{i+2} = \left(\sqrt{D_i} + \sqrt{D_2}\right)^2 \\
\Delta_{i+2} = 2\sqrt{D_iD_2}
\]  

(D.8)

This may be generalised by using a vectorial notation along with a measure of coupling (analogous to covariance) between the sections mediated by the common traffic:

\[
D_{i+2} = \left(\sqrt{D_i} + \sqrt{D_2}\right)\left(\sqrt{D_1} + \sqrt{D_2}\right) = D_1 + D_2 + 2c_{12}\sqrt{D_1D_2}
\]

(D.9)

As pointed out in section D.1, this delay is potential and will be realised only to the extent that the speed of the traffic in the queues is less than its normal free-flow speed. The speed factors applying to the two sections are in general different.

If \(p_{12}\) is the proportion of the section 1 queue that continues onto section 2, the proportion of section 2 queue drawn from section 1 is considered to be given by the ‘transposed’ proportion:

\[
p_{21} = \frac{q_1}{q_2} \quad (D.10)
\]

Then, in outline, according to the two-section model, if \(f_i\) are the section speed-dependent factors and \(\beta\) is the apportioning factor, itself a complicated function of the section characteristics, total delay is:

\[
D_{i+2} = D_1 + D_2 + 2(f_1p_{12}\beta + f_2p_{21}(1-\beta))\sqrt{D_1D_2}
\]

(D.11)

This additive form is questionable as there seems no reason why \(D_2\) should enter in full if continuing section 1 traffic forms only a fraction of section 2 traffic. The problem is that with the different speed factors, it is not possible to average the original and transposed proportions. The method is also so complex that its extension to more than two sections appears impractical. However, the extension of (D.9) to more than two sections appears straightforward, for example with three consecutive sections:

\[
D_{i+2+3} = \left(\sqrt{D_i} + \sqrt{D_2} + \sqrt{D_3}\right)\left(\sqrt{D_1} + \sqrt{D_2} + \sqrt{D_3}\right) = D_1 + D_2 + D_3 + 2c_{12}\sqrt{D_1D_2} + 2c_{23}\sqrt{D_2D_3} + 2c_{13}\sqrt{D_1D_3}
\]

(D.12)

If \(c_{12}=1\) and \(c_{23}=1\), ie all queuing traffic passes across both junctions, then the last three terms will be maximised, so clearly \(c_{13}\) must be zero. Hence it seems reasonable to assume that the \(c_{ij}\) are closely related to the \(p_{ij}\) for all relevant \(ij\) pairs.

We see in Figure D.1 that the quicker downstream the tail of the queue moves, the more delay is apportioned to later sections. Conversely, once the tail moves upstream extra delay is allocated to the first or upstream sections. In the special or ideal case where the tail of the queue does not move at all, as in Figure D.2, the problem of apportionment does not arise. In this case, there is similarity of shape respectively between the primary delay regions, and between the extra delay regions, and each belongs to (ie is incurred upon) the first section with which it is associated, so for each term:

\[
2c_{ij}\sqrt{D_iD_j} \quad \rightarrow \quad \text{potential delay to section } i \quad (\forall j)
\]

(D.13)
Assuming that the discharge wave in on each section (ie the right-hand bold boundary of each primary queuing region in Figure D.2) represents the divide between the primary and extra queues, although in practice its speed may be different, each extra queuing region \( ij \) appears to have the queuing characteristics of the determining section, \( j \), rather than the incurring section, \( i \). Nevertheless, the queue speed should be compared with the normal speed of traffic on the \textit{incurring} section \( i \). This leads to a separate speed factor for each extra queue region, removing the averaging problem noted earlier:

\[
f_{ij} = \left(1 - \frac{v_{bj}}{v_{ai}}\right)
\]

When we consider non-adjacent links, \( p_{ij} \) is the proportion of traffic on \( i \) which is affected by the queue developed on \( j \), but it relates entirely to traffic on section \( i \), so assuming that traffic is homogenous on each section:

\[
p_{ij} = \prod_{m=i}^{m=j-1} p_{mm+1} \quad (j>i+1) \tag{D.15}
\]

Also, we define:

\[
p_{ij} = 0.5 \quad (\forall i, \text{ for consistency})
\]

\[
p_{ji} = p_{ij} \frac{q_j}{q_j} \quad (j>i+1) \tag{D.16}
\]

finally assuming a symmetrical application of the proportions:

\[
2f_{ij} \sqrt{p_{ij}D_i \sqrt{p_{ji}D_j}} \rightarrow \text{actual delay to section } i \quad (\forall j) \tag{D.17}
\]

While this model can accommodate multiple sections, it pushes the limits of respectability, depending heavily of assumptions of symmetry and complicated parameters which have a matrix nature. This is not really compatible with a 'section-based' approach.
APPENDIX E – QUADRILATERAL MODEL OF LINKED SECTION QUEUES

E.1 Quadrilateral queuing regions

In order to pursue a section-based approach which is as realistic as possible when queues extend over more than one road section, we return to looking at what happens physically at the boundary between queue sections. In this model, the denser than normal discharge flow from a queue on an upstream section modifies the input flow and hence the queue growth on the section downstream, as diagrammed in Figure E.1, using the notation and variable names adopted earlier in the Report, with the addition of unbracketed subscripts referring to the section (1 or 2) and bracketed subscripts referring to the queuing phase (queue carried over, or free-flowing arrivals).

![Figure E.1 Several possible configurations of the queuing region](image-url)
The flow across the ‘stop line’ ahead of the queue goes through periods of normal flow $q_a$, reduced flow $q_c$ around the obstruction, and increased flow $q_d$ as the queue discharges. This must occur if the queue dissipates in a finite time, to conserve the total volume entering the section during period $t_a$. In the current model the wave speed of discharge to free flow $v_d$ is fixed at $\lambda/\tau$, assumed in calculations to be -20 km/h. The discharge flow $q_d$ in this case is assumed to be equal to the carriageway capacity, although it is believed that in reality it should be a bit less\(^{18}\).

The duration of increased flow as defined earlier in Section B.6, as measured at the intersection is easily calculated from the geometry of Figure E.1:

$$t_f = t_d \left(1 - \frac{v_{bd}}{v_{d1/b1}}\right)$$

(E.1)

If the tail or change wave speed is $v_{ab(1)^r}$, an adjustment needs to be made allowing for the traffic speed to determine the time at which normal demand flow $q_{a2}$ takes over, representing phase 2. This time relative to the start of movement on section 2 is:

$$t_{f(1)} = t_{f1} \left(1 - \frac{v_{ab(1)}}{v_{d1/b1}}\right)$$  subject to maximum of  \( t_{f(1)} + \frac{H}{v_{d1/b1}} \)

(E.2)

The internal and discharge properties of the queue on section 2, which depend only on passing capacity and obstruction speed, are not altered, but the speed of the tail of the queue changes, as a result of enhanced net demand on the section, as shown in the cases at the top of Figure E.1.

The possibility exists that the phase of increased flow, lasting until $t_{f(1)^r}$, continues beyond the point where the queue on section 2 has discharged. In this case the queue shape is triangular again, and in effect we have $v_{ab(2)}$ equal to $v_{ab(1)^r}$. In principle, the excess heavy flow ought to be projected onto the next sections downstream, but this could become very complicated, and it is likely to occur in reality only when $v_{ab(1)}$ is positive (downstream) and section 2 on which it impinges is short.

Conversely, the situation can occur where a queue develops as a result of an initial heavy inflow where no queue would develop because of the normal flow. The second ‘tail wave’ then advances quicker than the load and may even catch it up (see two top left cases in Figure E.1). The simple test for queue presence, equation (7), no longer applies, but equation (9) remains valid.

**Considering section 2 only and dropping subscripts 1 and 2**, the time $t_e$ at which the tail wave intersects the discharge wave can be calculated by geometry:

$$t_e = \frac{(v_L - v_{bd})f + (v_{ab(2)} - v_{ab(1)})f_{f(1)}}{v_{ab(2)} - v_{bd}}$$

(E.3a)

If this results in $t_e < t$ then recalculate as:

$$t_e = \frac{(v_{ab(2)} - v_{ab(1)})f_{f(1)}}{v_{ab(2)} - v_L}$$

(E.3b)

Otherwise if $t_e < t_{f(1)}$ then recalculate as:

$$t_e = \frac{(v_L - v_{bd})f}{v_{ab(1)} - v_{bd}}$$

(E.3c)

---

\(^{18}\) Personal communication by Tim Rees.
The maximum queue extent is the greater of the $X(i)$ shown in the Figures E.1, depending on whether $t_f(1)$ is less or greater than $t$. If we write:

$$t(1) = \min(t_f(1), t)$$

$$t(2) = \max(t_f(1), t)$$  \hspace{1cm} \text{(E.4)}

then

$$X(1) = t(1) \left( v_L - v_{ab(1)} \right)$$  \hspace{1cm} \text{(E.5)}

$$X(2) = (t_e - t(2)) \left( v_{ab(2)} - v_{bd} \right)$$ \hspace{1cm} \text{if} \hspace{0.5cm} t_e > t_2  \hspace{1cm} \text{(E.6)}

$$X = \max\{X(1), X(2)\}$$  \hspace{1cm} \text{(E.7)}

The $X$s must of course be restricted to non-negative values. If the queue is assumed to be homogeneous, the maximum queue size in terms of vehicles can then be calculated as:

$$L = X k_b$$  \hspace{1cm} \text{(E.8)}

The total delay in the queue, again assuming it to be homogeneous, is proportional to the queued area, adjusted according to the speed of the traffic. This is equal to the sum of the areas of the left-hand and right-hand triangles, and the part-parallelogram between them. Hence, where ‘(u)’ stands for ‘unadjusted’, and the speed adjustment assumes the regions concerned have similar free-flow speeds:

$$D_{b(u)} = .5 k_b \left( t(1)X(1) + (t_e - t(2))X(2) + (t(2) - t(1)) \left( X(1) + X(2) \right) \right)$$

$$= .5 k_b \left( t(2)X(1) + (t_e - t(1))X(2) \right)$$  \hspace{1cm} \text{(E.9)}

$$D_b = D_{b(u)} \left( 1 - \frac{v_b}{v_u} \right)$$  \hspace{1cm} \text{(E.10)}

**E.2 Assessing total queuing**

Figure E.2 is an illustrative example of a multi-section queue. The total ‘area’ of queuing is well-defined and the total delay can be calculated as the sum of the delays ascribed to (induced on) the individual sections. The model deals with the three quadrilateral and one triangular queuing regions which together make up the total queuing area, each being associated with one road section. It can be seen that some delay is not incurred on the section on which it is induced, but apportioning delay between sections is impractical and largely irrelevant since estimation of delay from actual data, eg MIDAS, is likely to be done for whole queuing regions.

However, it is often desirable to make an estimate of the whole queue which might be observed over the whole region. Figure E.2 shows that queues can be measured in several incompatible ways, so the ‘overall queue size’ is no easy to define. The maximum queues induced on each of the sections, as defined in Figure E.1 and calculated by equation (E.7), cannot be added together since they occur at different times, and can leave gaps or even overlap. The question is, how can section-based queue information, allowing explicit relationships only between adjacent sections, be used to estimate queue size over multiple sections? A case in point is where a queue forms on $n$ adjacent identical motorway sections under constant demand following a moving load. If the final queue size on the first section is $L$ then the final queue over all the sections will be $nL$. However, the maximum section queue will never be much greater than about $2L$. It would be useful to have an estimate of ‘linked’ queue size which behaves like the first case, but clearly this has to be synthesised since it cannot be extracted directly from the geometry of section queues.
Figure E.2  Arbitrary queue caused by load moving on four consecutive sections at varying speed, showing complex shape of queuing regions and vehicle paths, and variability of queue measurements.

Queue extents determined by each section calculation could exceed the physical size of the section, but the actual queue which is observed on any section is limited to its physical length. The remainder can be ascribed to the section upstream, but parallels rather than augments any queue already belonging to that section. The following adjustment to queue extent adjusts the ‘local’ queue for any tailback from the section downstream, which is considered to exist in parallel to it: if the maximum section-based queue $X_i$ is as calculated by (E.7) for each section $i$ of length $H_i$, we define the queue adjusted for tailback recursively as:

$$X_i^\triangleright = \max(X_i, \max(X_{i+1}^\triangleright - H_{i+1}, 0) \cdot \frac{\pi_{i+1}}{\kappa_{i+1}})$$  \hspace{1cm} (E.11)

where:

$$\pi_{i,i+1} = p_{i,i+1} \left( \frac{q_i}{q_{i+1}} \right) = \text{proportion of average } q_{i+1} \text{ originating from section } i$$

$$\kappa_{i,i+1} = \frac{k_i}{k_{i+1}} = \text{density factor from } i \text{ to } i+1 \text{ in a queue at constant volume}$$

(E.12)

The first factor reflects the fact that only part of the tailback from section $i+1$ contributes to queuing on section $i$, the remainder being ‘off route’, depending on the turning movements between the sections. It is complicated by the ratio of average volumes, which depends on flows which can differ.
during the two phases in the quadrilateral model, though it is unlikely to be much different from the value that would be obtained from using normal demand volumes. The second factor allows for change in geometry (principally number of lanes) as the queue is transferred from section \( i+1 \) to section \( i \). Using equations (2), by equating expressions for flow in terms of density and section geometric parameters, the ratio of densities can be obtained, subject to a minimum density on \( i \) equivalent to that at capacity. Note that when queuing traffic moves from fewer lanes to more lanes, the flow is sustained at lower speed and higher density, so the \( \kappa \) factor is less than 1. Since it acts as a divisor in (E.11) this has the effect of increasing the extent of the queue on the narrower section.

We now define a notional ‘linked queue extent’ recursively in the forward direction by:

\[
X_{i+1}^+ = X_{i+1}^{\triangleright} + p_{i,i+1} \left( 1 - \frac{t_{i+1}}{t_{d,i}} \right) X_{i}^+ \tag{E.13}
\]

The adjustment of queue size in vehicles proceeds similarly, first defining (E.14) then calculating \( \{ L_i^\triangleright \} \) by the analogue of (E.13):

\[
L_i^\triangleright = k_{bi} X_i^\triangleright \tag{E.14}
\]

### E.3 Example of multi-section results

A 3-lane motorway has five sections each of 2.5 miles length, with normal demand of 3200 veh/h and a PCU factor of 1.1. The load moves at 9 mph along these sections, occupies 2 lanes, and leaves a passing capacity of 1150 veh/h at 30 mph. Figure E.3 shows the total delay on all the sections depending on how many have linked queues (ranging from none to all, working downstream from the first) and the following factor between linked sections.

![Figure E.3 Effect of linking motorway sections using quadrilateral model](image)

It can be seen that the effect of linking is cumulative. The results include passing delay, and there is a minimum delay of about 99.6 veh-h on each independent section, but for a following factor of 1 the
delay goes up roughly as the square of the number of linked sections, as expected. It also appears to go up by a power of the following factor, so the extra delay with a following factor of 0.5 is about a quarter of that with full linking. This is because the effect of traffic ‘leaking away’ is cumulative.

E.4 What following factors should be used?

If as assumed above two road sections have the same normal flow, then the flows entering and leaving their junction must be the same. Does this not mean automatically that they are 100% linked? In fact the flow out of a queue can be much greater than the normal free flow. This difference between the demand components in equation (21) means that the demand on the downstream section will depend on the proportion \( p \) even if the normal volumes on the sections are identical. In fact, nothing can be inferred about \( p \) from the normal volumes, so the following factors must be determined empirically.

If default values for turning proportions at junctions can be found, they will depend upon the road types involved. For typical route models, the road sections may consist of more than one junction-to-junction section itself. In these cases, some ‘leakage’ of queuing traffic may be expected within the route section. As there is no provision for this in the queuing model, it must be allowed for by reducing the following factor, and possibly by adjusting the average passing capacity over the section.

The most logical way to do the former is as follows:

\[
P_{\text{route-section}} = p_{\text{between-junctions}} \times \max\left(\frac{H_{\text{route-section}}}{H_{\text{average-between-junctions}}}, 1\right) \tag{E.15}
\]

E.5 Incorporating breaks

Although the principles described in Appendix C still apply, the critical role of the onward flow to the next road section means that the effect of breaks must be fully integrated into the model. In Figure E.4 (left), it can be seen that the effect of even a single centrally timed break on the quadrilateral queue is very complicated, and it would be impractical to attempt to model this exactly even in this case, let alone for an arbitrary sequence of breaks.

![Quadrilateral queue with break](left), and triangular approximation (right)
The simplified triangular analogue in Figure E.4 (right) starts by defining a triangular queue with the same area (hence delay) as the original queue. Because of the shift in the point of final dissipation, the effective tail-wave speed \( v_{ab} \) is not the same as that produced by the combined effect of \( v_{ab(1)} \) and \( v_{ab(2)} \). Also noticeable is that a different shift in the end time of the queue \( t_e \) occurs in addition to that resulting from break time \( t_B \). Finally, it is not obvious that the approximate area \( Q_1+Q_2 \) is equal to the ‘true’ queuing area \( Q_1+Q_2 \). Probably the most unsatisfactory feature of the approximation is the length of the discharge period of the second (post break) queue, since this appears to be particularly sensitive to the shape of the queue and the placing of breaks. Any traffic already discharged and so ahead of the load can be assumed to produce no extra delay since it has available the full capacity of the road.

However, the rate of discharge flow is not altered by breaks, so their effect can probably be covered adequately by recalculating the proportion of the total queue following the load. Thus the effect of breaks is accounted for in practice by adjusting the proportion of the total discharge flow added to the demand on the next section, \( p \) in equation (21), by the same factor as that applied in (C.6).

E.6 Qualifications

The following points about the quadrilateral queuing model need to be kept in mind:

- Queue lengths on different sections, whether their values are ‘linked’ (ie including spillback) or not, are evaluated at different times, so where a queue persists from one section to the next the modelled values do not represent a single ‘snapshot’. The method of linking queue lengths can compensate for this to an extent;

- The model as a whole breaks down when ambient demand exceeds free-flow capacity. When demand approaches capacity, apparently strange things can happen. For example, the mechanism for passing on queues works on the excess of flow on the upstream section over the assumed ambient traffic there. Since queuing flow tends to be close to but not exceeding the section capacity, if the ambient flow is close to capacity, the excess flow may be very small, giving the impression that queue has ‘disappeared’ even if the following factor is high. In fact the flow arriving at the intersection is hardly different from ‘normal’, but the following factor is now probably inconsistent with the normal traffic situation as represented by the ambient flows specified for the sections. Unrealistic assumptions will therefore lead to unrealistic results.

- The model makes simplifying assumptions about the traffic state in the queue induced by each road section. In particular, if the queue propagates onto sections upstream, no account is taken in the basic model of the effect of change in road geometry which, although it may not affect flow, will result in changes in speed and density. The problem is, if such an adjustment is made, how far ought it to go? Exact treatment could involve backtracking over many sections, which is too complicated for a section-by-section model, but if only the geometry of the first upstream section is considered, this could be misleading if earlier sections have quite different geometry.

---

19 This also applies when altered demand incorporating queuing from upstream and following factor exceeds the downstream section capacity. However, this can be allowed for by lengthening the phase (1) duration \( f(t,1) \) and reducing the flow correspondingly.
E.7 Calculation of average delay per vehicle

An approach which avoids counterintuitive results, as alluded to earlier in Section B.6, is based on a slightly different way of looking at queues segments compared to the previous model. In the familiar quadrilateral queue segment shown in Figure E.5, instead of considering traffic entering the queue segment, which involves the complex shape of the queue tail, we consider traffic leaving the queue. In the case shown, the speed in the queue, $v_b$, is supposed to be greater than the speed of the load, $v_L$, consistent with some traffic being able to pass (otherwise the speeds must be equal). The total vehicles which are affected by the queue can be separated into three parts:

![Figure E.5 Separation of flows departing from a queue segment](image)

(A) The Ambient flow not affected by the queue (not involved in the calculation).

(B) The flow remaining Behind the Load up to the end of the section:

$$Q_B = \left( 1 - \frac{v_{bd}}{v_b} \right) q_b (t_e - t)$$

(E.16)

(C) The flow which escapes the Load through the passing Channel:

$$Q_C = \left( 1 - \frac{v_L}{v_b} \right) q_b t$$

(E.17)

(C) The queue flow $Q_b$ divides further according to whether it Exits the route at the end of the section or Follows the Load onto the next section:

$$Q_{BE} = (1 - p)Q_B$$

$$Q_{BF} = pQ_B$$

(E.18)

(E.19)
The speed dependent factors in (E.16) and (E.17) arise from geometry when calculating the duration of the lines SP and QR in Figure E.5, representing horizons passed through by traffic moving at speed $v_b$, destined either to pass the Load (left) or to remain in the queue until the end of the link (right).

When looking at a single segment of the queue in isolation, which in the model is treated as proxy for the effect of a single road section, it is arguable that all the flows should be included, so the delay per vehicle is given by (where $i$ signifies the section):

$$d_{b(i)} = \frac{D_{b(i)}}{Q_{B(i)} + Q_{C(i)}}$$  \hspace{1cm} (E.20)

However, when looking at the route as a whole this leads to multiple counting, and is replaced by:

$$d_{b} = \frac{D_B}{\sum_{j=1}^{N} (Q_{BE(j)} + Q_{C(j)}) + Q_{BF(N)}}$$  \hspace{1cm} (E.21)

The inclusion of the last term in the denominator is actually superfluous if the following factor for the last route section, $p_{N}$, is set to zero, but the model can be specified recursively in such a way that the result is independent of this factor.

Equation (E.20) is not entirely satisfactory as a description of the effect of the queue on a single road section. The quadrilateral queue model, developed to keep the amount of calculation linear with the number of road sections, can spread each queue segment across several road sections, where the queue is present at different times. To calculate delay to an individual vehicle between entering and leaving a single road section, and by implication any subset of road sections, it is necessary to divide a multi-section queue ‘horizontally’ (refer to Figure D.2 earlier), but this means each slice of the queue depends on several road sections – in principle an unlimited number, which is what the present model is designed to avoid. Accurate estimation of delay for vehicles on specific paths intersecting the load’s route at specific times may therefore require a more sophisticated model.
APPENDIX F – ESTIMATING DEMANDS AND CAPACITIES

The model is principally sensitive to the demand, and the capacity for passing the load, and to a lesser degree the speed of the load, on each road section. These quantities are relatively easy to measure, at least over short periods, but capacity passing the load is difficult to estimate because the driver behaviour which determines it is not understood.

F.1 Sources of demand data

Demand varies significantly with day of week and time of day, as well having a random element. Many abnormal loads, especially the larger, heavier ones, are scheduled to travel at specific times on specific days, so for delay and cost calculations to be meaningful it is essential that realistic time-dependent, not average, demand values be used.

For a limited number of road sections, recent historical data can be obtained from MIDAS records or the TRADS database. MIDAS resolves flows to one minute for specific days, while TRADS flows are resolved to one hour and aggregated for different day types. For other roads only Average Daily Traffic (ADT) or Average Hourly Traffic (AHT) flows collected in a single survey period may available, and failing that just typical values for the type of road. The demand at the time the load is likely to travel must then be estimated using a time-dependent profile appropriate to the road type and day of travel. DMRB Vol 12 contains an illustrative weekday Automatic Traffic Counter profile with finer time resolution but also divided between four flow groups (broadly: overnight, AM peak, inter-peak and PM peak). However this profile applies only to weekdays.

The QUADRO Manual (DMRB Vol 14) describes a way of estimating a daily profile from average daily or hourly flow for each of several day types. Although these profiles are specified on an hourly basis, they are defined in terms of only four flow groups in any one day, so have limited resolution. By interpolating them, a spreadsheet-based Profile Calculator has been constructed which enables flow to be estimated for any period in any normal day. Figure F.1 compares plots from these three sources, though not for equivalent roads. It can be seen that the onset of the peak is later and gentler on Saturday and Sunday mornings than on weekdays, which is why loads which have to move in daylight are usually scheduled then.

F.2 Capacity passing a load

Capacity passing a load on a given number of lanes is believed to be reduced compared to free-flow capacity. Roberts et al (1994) showed that capacity past an incident is reduced on average by 24.2%. ITEA has adopted a capacity factor of around 0.8 based on this. Capacity might be reduced by drivers taking extra care, or by simply slowing down since maximum flow (on motorways and similar) occurs at relatively high speed. The flow possible when passing a vehicle moving at moderate speed could in principle be higher than that possible when passing the confusion and stationary obstacles associated with an incident.

However, measured average flow using one lane, past a wide slow load occupying two lanes of the M6 on a Saturday, was 1396 veh/h/lane (sample size 15 minutes), compared to about 2000 veh/h/lane which would be expected in free flow based on a ‘standard’ 2300 PCU/h/lane and 12% heavy vehicles. The speed in the passing lane was estimated to be about 30 mph, around half the normal speed at maximum flow of 65-67 mph. It is virtually impossible to determine whether this speed drop was a result of drivers specifically reducing their speed or extending their time headways, because the two are linked through the coupled speed/flow relationship. Because of non-linearity (see Figure 4 in Appendix B) a drop in flow of 20-25% is consistent with more than halving speed and vice-versa.

---

20 HGV and PSV percentages vary by road type, day of week and time period in the day. PCU factor is in the range 2.0-2.5. If practical, a range of appropriate default values will be assembled for the final model implementation.
Figure F.1. Comparison of DMRB Vol 12 typical weekday profile (top), sample of TRADS average count profiles from the M6 (middle), and interpolated QUADRO profiles (lower), where colour of graph codes day of week: red=weekday, dark blue=Saturday, magenta=Sunday
What actually determines capacity? Flow is simply the inverse of the average time headway between vehicles, so maximum flow, as measured at a fixed point, is the inverse of the minimum time headway. The model of coupled flow defined in Appendix B contains two parameters, a minimum distance headway $\lambda$ and a minimum time headway $\tau$ which are assumed practically, though not necessarily, to be independent of speed. Distance headway dominates only at low speeds. The maximum flow is therefore just $1/\tau$, although this is never achieved in practice. Calibration of the model is made easier by assuming that the speed of shock waves, whose magnitude is $\lambda/\tau$, is constant at around 20 km/h in the upstream direction, as is approximately true in practice.

The model is then calibrated by specifying the maximum practical flow and the speed at which it occurs in the absence of the load, both of which are available from measurements or from standard free-flow relationships. The maximum speed in the coupled regime is effectively constrained by the free-flow relationship, which itself places little restriction on the maximum flow. The free-flow relationship arises from interactions between vehicles which are different from those of the coupled regime, and also from physical or legal limitations on speed.

Traffic constrained by the capacity of the passing channel is likely to be governed by a coupled regime similar to that applying to a normal speed/flow relationship, but the parameters could be substantially different because of the special circumstances or geometry. Additional driver caution or distraction might increase the time and distance headway parameters, reducing the flow at each speed in proportion. In practice, the major constraint is likely to be the actual speed drivers achieve in the channel, possibly ‘forced’ by the presence of police or related statistically to the speed of the load. Figure F.2 illustrates these various effects.

![Figure F.2. Ways in which reduction of capacity of a passing channel could occur](image-url)
APPENDIX G – MEASUREMENTS FROM WITHIN A CONVOY

G.1 Dunn’s formula

Dunn (1967) quoted a formula for calculating directly the delay to other road users in vehicle-hours/mile travelled caused by an abnormal load on a single carriageway road – ie not allowing any vehicles to pass.

\[
Delay = n \left( \frac{1}{V_2} - \frac{1}{V_1} \right)
\]  

(G.1 or D1)

In this, \(n\) is the mean number of vehicles affected, \(V_1\) is the mean speed (mph) of the vehicles under normal conditions, and \(V_2\) is the reduced mean vehicle speed caused by the presence of the load. In fact this formula contains the very simple message that delay is just the difference between travel time with and without the load, and rather disguises the fact that estimating these is not a trivial exercise if the queue is growing, whereas direct measurement is somewhat more straightforward. Measuring the quantities in the formula is also not easy if some traffic is passing the load or if observers cannot be mustered in the numbers Dunn had at his disposal (one every 2 miles). So this formula is not helpful for estimation purposes.

G.2 Delay in the Zone of Influence

In addition to a queue which builds up and decays, traffic can incur delay as a result of reduction in speed while passing the load. The co-moving length over which this occurs is the ‘Zone of Influence’. In practice, especially over short periods of time, it may be difficult to tell the difference between a queue caused by insufficient capacity and vehicles waiting to pass in the Zone of Influence, particularly as the latter could change if the speed of the load changes.

Although Dunn did not use the term ‘Zone of Influence’, he estimated delay in terms of the ‘average number of vehicles in the queue’ or ‘following vehicles’ \(n\). This translates into our terminology as follows: the total delay to an average of \(n\) vehicles in a queue whose normal speed is \(v_a\) affected by an abnormal load moving at speed \(v_L\) for a distance \(H\) is given by:

\[
\Delta = nH \left( \frac{1}{v_L} - \frac{1}{v_a} \right)
\]  

(G.2)

If a queue is not actually growing, vehicles obstructed by a load may be obliged to travel through the Zone of Influence at a reduced speed \(v_c\), which is still greater than the speed of the load. The length of the Zone of Influence, \(Z\), depends not just on the length of the load, but also the position of the escorts, the number of vehicles affected and their density \(k\), which in turn depends on their speed (see Figure G.1). The effective Zone could also be extended by random queuing effects, where random variations in the arrival rate and capacity cause a net queue whose lengths tends to a constant value. For these reasons, it is probably best determined empirically, although a value in the range 0.1 miles (for a faster self-escorted load) to 0.5 miles (for a slow police-escorted load) is typical.

Dunn treats the average number of vehicles in the queue behind the load \(n\), and the total number of vehicles affected \(N\), as synonymous. Clearly, this is wrong if vehicles are able to overtake. Equation (G.2) should therefore be replaced by:

\[
\Delta = Nz \left( \frac{1}{v_c} - \frac{1}{v_a} \right)
\]  

(G.3)
Figure G.1. Vehicles delayed in the Zone of Influence of an Abnormal Load

where $Z'$ is the effective Zone of Influence seen by a passing vehicle, i.e., the distance travelled at speed $v_c$. This is greater than $Z$ because of the movement of the load and, by geometry, is given by:

$$Z' = Z \left(1 - \frac{v_c}{v_L}\right) \quad \text{(G.4)}$$

Thus:

$$\Delta = NZ \frac{(v_a - v_c)}{v_a(v_c - v_L)} \quad \text{(G.5)}$$

Formulae (G3-4) are no longer valid if $v_c = v_L$, but in that case there is no real distinction between the queue and the Zone of Influence, which grows with time as more vehicles are caught up in the convoy. A similar situation occurs if the passing capacity is insufficient for the demand, except that the speed in the queue ($v_b$) will generally differ from that in the passing traffic which is otherwise free flowing, and a distinct Zone of Influence in the sense of Figure G.1 still exists. In these cases it is better to use the queuing model (see earlier in Appendix B) and treat the queuing delay as additional to the delay incurred while passing through the Zone of Influence. Another possibility is that a few vehicles, in particular HGVs, may decide to stay behind the load, and so incur greater delay, while others pass. To deal with these various cases and possible ambiguities, $NZ'$ in (G.2) may be considered to be a single variable: the estimate of the total vehicle-kilometres affected by the load as it covers the section length $H$. 

G.2 Estimation of delay from values measured en route

When travelling with an Abnormal Load, it can be difficult to distinguish between a queue and the Zone of Influence, and it is virtually impossible to measure a queue of more than about 20 vehicles from the viewpoint of an escort vehicle. It is also difficult to tell how fast vehicles in the queue are moving, and how fast vehicles would travel if the load were not present. However, it will usually be possible, over a short section or period, to note the speed of the load, the speed of passing traffic, the number of vehicles which passed and their speed, and the number of vehicles in the final queue, provided it is not too large or growing.

The time taken by a queue to discharge, which contributes to $NZ'$, is difficult to observe, though this can be compensated for to a degree by averaging initial and final queue lengths so a queue contributes to delay on two sections. A simple estimate of total delay is:

$$\Delta = \left(\frac{L_0 + L_H H}{2v_L} + \frac{N_c Z}{v_c - v_L}\right) \left(1 - \frac{v_c^*}{v_a}\right)$$

where $v_c^* = \max(v_c, v_L)$ (G.6)

where $L_0$ and $L_H$ are the initial and final queue sizes on the section length $H$ and $N_c$ is the number of vehicles which pass on the section (this of course being zero if $v_c$ is equal to $v_L$, or it is not possible for traffic to pass). The first part of (G.6) is simply the average queue length multiplied by the time taken by the load to traverse the section. The second part comes directly from (G.5), which already embodies the correction factor for the speed of passing/following traffic. Any delay to queues on the opposing carriageway or side roads is additional. This can be difficult to estimate from the convoy because the time for which the queue is present is not known, neither is the flow which fed it.

Where queues are substantial, direct observation from within the convoy is not reliable and modelling or observation from multiple stations is likely to be required.

G.3 Estimation of passing flow from observation

Passing flow is easily observed when travelling with a convoy, but what is important for modelling queuing is the passing flow which would have been measured by a stationary observer. This is in principle given by the formula:

$$q_{c(\text{static})} = \frac{q_{c(\text{moving})}}{1 - \frac{v_L}{v_c}}$$ (G.7)

However this formula is unreliable if the difference between the load and passing speeds is very small, as can often be the case, even while as measured by a stationary observer the flow in the passing channel remains substantial. According to equation (1) in Appendix B, passing flow could also be estimated by multiplying the passing speed $v_c$, which can be estimated to reasonable approximation, by the density of the traffic. However, density is difficult to measure because it is defined only over a finite distance, which has to be long enough to ‘iron out’ spatial variations, and this is difficult to observe. In conclusion, in some cases it may be difficult or impossible to assess the ‘static’ passing flow from within a convoy, so it will be necessary to rely upon values estimated according to the principles of Appendix F.
APPENDIX H – SOME MONITORING RESULTS

H.1 A real case example

On 15 and 22 May 2004 identical slow wide abnormal loads travelled from the Areva site in Stafford to Seaforth near Liverpool. Each consisted of a transformer carried on a ‘girder-frame’ trailer hauled by a locomotive at each end. Each vehicle had a gross weight of 294 tonnes, overall length of 60 metres, and width of 5.9 metres, and moved at a typical speed of about 12 mph except uphill. The journeys took place in two stages: on Saturday from Stafford via M6 Junction 14 to a reserved area at M6 Junction 16; then on Sunday continuing up the M6, then onto the M62 to Liverpool. Substantial queuing was observed behind the moving loads between J14 and 16 on the Saturdays. Figure H.1 shows the inputs and outputs of the spreadsheet model for the 22/23 May move, with estimated costs.

Figure H.1. Spreadsheet model of Stafford-Seafort M6/M62 AIL move on Sat/Sun 22/23 May

21 This represents the output of the Model at the time of reporting. Results of subsequent versions could differ. Channel capacities are not adjusted since the values are based on direct measurement.
MIDAS data are patchy between Junctions 14 and 16, thanks to a temporary lack of data network bandwidth, but there is sufficient coverage to enable the queue on the M6 on Saturday 22 May to be reconstructed from MTV plots. In Figure H.2, which combines two plots, queuing is visible as light (low speed) areas, and the primary queue can just be distinguished towards the bottom. The MTV plot has superimposed on it the estimated boundaries defining the main queue and discharging flow (white), and the trajectory of a typical vehicle (yellow). There is possible evidence of a ‘knock-on’ effect near Junction 15 produced by concentrated traffic discharging from the main queue.

Figure H.2. Annotated MTV plot of queue produced by AIL on M6, Saturday 22 May
The boundaries of the main queue are:

- (left) the load entering at J14 and moving to J16 at an average speed of 20.4 km/h (12.7 mph),
- (lower) the back of the queue moving downstream at around 7.7 km/h as the queue grows,
- (middle) the discharge wave at the front of the queue moving (upstream) at -18.6 km/h after the load has left the carriageway.

The spreadsheet model (working not shown in Figure H.1) estimates slightly different speeds for the queue boundaries: 8.2 and -20 km/h respectively, the latter being the value assumed for the ratio of distance headway to time headway. The value assumed for demand flow, 3350 veh/h, is the means of the range observed on 8/5 and is consistent with TRADS data for that month. The passing capacity is essentially a rounded version of that estimated from limited counts ahead of the load on 15/8.

It is assumed that the queue discharges at the maximum capacity of the carriageway. It is quite easy to show that the model of coupled traffic, equation (4) in Appendix B, which is also based on fixed distance and time headway parameters, then leads automatically to the speed of the discharge wave assumed. Should the capacity of the discharging queue front be less than the free-flow capacity, the speed of the discharge wave will be reduced. It can be argued that the reduction in capacity ought to be similar to that which occurs after flow breakdown (this effect could be held responsible for the self-sustaining character of flow breakdown). However, while the results are not inconsistent with this, it is felt that the model and measurements are not sensitive enough to demonstrate such an effect.

A similar but more intense version of the ‘knock-on’ pattern shown in Figure H.1 occurred at the same place after the 15 May move. It is hard to resist the conclusion that this was not just an isolated event, but a result of the high discharge flow, close to the carriageway capacity, hitting a critical point downstream. The shape of the secondary queuing region is not clear cut, but this may be generated by multiple seed points along the carriageway. The overall duration of the secondary queue can be presumed to be consistent with the Section B.6 model. This has optional allowance for dispersion of the discharging traffic, but there is little evidence that it is significant in this case.

The maximum queue length estimated from observation on 15 May was 14 miles. Estimates from MTV and Model are in good agreement at 13.5 and 14.1 miles respectively. Delay is inherently more difficult to measure, although a routine in MTV exists to do this, and the limited MIDAS coverage between Juctions 14 and 16 adds further complication. Using a reference speed of 65 mph, delays of 731 and 346 veh/h respectively were calculated for 15 and 22 May. From the plots, the proportion of the queuing region covered by loop data can be estimated as 19.6% and 17.5% respectively (one more loop provided data on 15/8). Factoring up to ‘fill in’ the unrecorded sites gives delays of 3739 and 1975 veh-h respectively, while the model estimates 4017 and 2669 veh-h respectively, including the short stop at Junction 16, but excluding any knock-on effects as these are not covered by MTV.

These results validate the queuing model in terms of traffic behaviour, and can be considered a reasonable validation in terms of delay values.
H.2 Indications from MIDAS analysis

Initial results from analysis of the MIDAS traces of about 60 moves indicates that the mileage distribution of delay costs, on the motorway sections where MIDAS is installed, is very extended (see Figure H.3), with no delay occurring on about half of all sections, and few miles with cost exceeding £100/mile. The median cost is £2.50/mile. There cannot be said to be a ‘typical’ cost-per-mile, but the average value over all loads works out at £55/mile. Figure H.4 provides another way of looking at this: delay caused per km on logarithmic scales. Although erratic, the relationship has the appearance of a ‘power law’ where there is no typical or modal value of delay.
APPENDIX I – ADDITIONAL FACTORS AND IMPACTS

I.1 Economic cost of delay

The current method of calculating the economic cost of delay is defined in WebTAG 2002, the DfT’s web-based Transport Analysis Guidance (http://www.webtag.org.uk, unit 3.5.6). Although it is necessary to assume an average traffic composition, costs could in principle be adjusted for road type, day of week and time period, all of which can affect traffic composition. However, the resulting variation in cost estimates is no more than about 10%, whereas the uncertainties in cost estimates resulting from uncertainties and variability in basic inputs like traffic volumes are likely to be of a similar order, while the assessment of cost-per-mile, say in comparison with a water alternative, is likely to depend on its order of magnitude. Consequently, it is felt sufficient to adopt DfT’s recommended average value of time of £11.28/hour (2002 prices).

The average proportion of HGVs on motorways is around 12%. Late at night, HGVs can compose a much higher proportion of traffic than during the day, but this is compensated to some extent by the fact that HGVs have a lower than average marginal value of time, around £8.50/hour. Also, late-night moves tend to have a low delay cost, so adjusting the value of time could have little effect on decisions about the move. According to stated preference studies, an increase in the variability of travel time can incur a cost additional to that of travel time. However, it is unclear how this should apply to the effect of abnormal loads, which are one-off events.

I.2 Impacts other than delay

Some factors have been listed in Section 5 of the main report. Their values are difficult to estimate, but this problem can be avoided to an extent if it can be shown that (apart from delay) these externalities are small enough not to affect major decisions in relation to Abnormal Loads.

Three sources provide some data (the first two included in the Literature Review):

- Surface Transport Costs and Charges Great Britain 1998 (Sansom et al 2001), which tabulates fully allocated and marginal costs of disaggregated road vehicle types in p/vehicle-km.
- A report by the Flemish Institution for Technological Research (VITO) (2004), which tabulates cost in €/1000 tonne-km from three sources for lorries, barges and trains22.

Converting between the different units is problematic. However, the factor for converting pence/lorry-mile to Euro/1000 tonne-km is simply 10/lorry weight, so no conversion is required if an average lorry weight of 10 tonnes is assumed. The data in Table I.1 can then be compared directly. It is questionable to identify as lorry-tonne with one AIL-tonne since these vehicle types affect the road and traffic in different ways. However, if we assume ordinary freight and AIL loads are comparable, and typical AIL weight is assumed to be 100 tonnes23, ie ten average HGVs, then these figures can be converted to £/mile values approximately by dividing by 10.

22 We are indebted to Mike Elsom of Sea & Water for bringing this report to our attention.
23 Actual statistics for loads between mid-2003 and mid-2004 are SO average weight 98 tonnes per move, VR1 weight 72 tonnes per authorisation (number of moves not specified) – see Final Report.
Subject to that qualification, the estimated cost of externalities is comparable with the median congestion cost found by MIDAS monitoring (Section H.2), while the most disruptive AIL moves cause congestion costs two orders of magnitude higher. This suggests that, where significant delay occurs, externality costs are unlikely to influence decisions about routing or mode.

---

From Surface Transport Costs and Charges Great Britain 1998 (Sansom et al 2001), Tables 7.4 and 7.5

Geometric mean of low and high estimates, 53% HGV-rigid, 47% HGV-articulated, converted to p/mile

HM Treasury GDP Deflator to convert to 2004 prices: factor 1.16

Original figures are tabulated for 1995 and for 2010 forecast. These have been interpolated pro rata.

Sources used by this report: VITO, EC and Planco tabulated individually

STC&C figures increased by 31% approximately, based on DEFRA recommendation for carbon costs

Vehicle operating and depreciation not separated

Upstream processes include emissions and infrastructure costs from vehicle and fuel production

Low and High estimates are respectively 31% below and 77% above this central figure