Estimating origin to destination travel by automatically reading licence plates

Prepared for Integrated Transport Economics and Appraisal (ITEA) Division, Department for Transport

B A Frith and T Sutch
This report has been produced by TRL Limited, under/as part of a contract placed by the Department for Transport. Any views expressed in it are not necessarily those of the Department.

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.
### CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive Summary</td>
<td>1</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>3</td>
</tr>
<tr>
<td>1.1 Objectives</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Background</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Research approach</td>
<td>4</td>
</tr>
<tr>
<td>1.4 Report structure</td>
<td>4</td>
</tr>
<tr>
<td><strong>2 Initial study</strong></td>
<td>5</td>
</tr>
<tr>
<td>2.1 Video recording</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Licence plate extraction from video</td>
<td>5</td>
</tr>
<tr>
<td>2.3 Initial analysis of manual and ANPR data</td>
<td>5</td>
</tr>
<tr>
<td>2.4 Licence plate validation</td>
<td>6</td>
</tr>
<tr>
<td>2.5 Results of correction and validation</td>
<td>7</td>
</tr>
<tr>
<td><strong>3 Relationship with determining factors</strong></td>
<td>10</td>
</tr>
<tr>
<td>3.1 Determining factors</td>
<td>10</td>
</tr>
<tr>
<td>3.2 Methodology</td>
<td>10</td>
</tr>
<tr>
<td>3.3 Results of analysis</td>
<td>10</td>
</tr>
<tr>
<td><strong>4 Model calibration and validation</strong></td>
<td>12</td>
</tr>
<tr>
<td>4.1 Proposed model</td>
<td>12</td>
</tr>
<tr>
<td>4.2 Model calibration</td>
<td>12</td>
</tr>
<tr>
<td>4.3 Model validation</td>
<td>12</td>
</tr>
<tr>
<td><strong>5 Discussion</strong></td>
<td>13</td>
</tr>
<tr>
<td><strong>6 Conclusions</strong></td>
<td>14</td>
</tr>
<tr>
<td><strong>7 Acknowledgements</strong></td>
<td>14</td>
</tr>
<tr>
<td>Appendix A: Using Automatic Number Plate Recognition (ANPR) equipment to estimate Origin to Destination (O/D) throughput using all-digit matching</td>
<td>15</td>
</tr>
<tr>
<td>Appendix B: Derivation of equations</td>
<td>19</td>
</tr>
<tr>
<td>Appendix C: Statistical theory</td>
<td>20</td>
</tr>
<tr>
<td>Appendix D: Fuzzy matching</td>
<td>21</td>
</tr>
<tr>
<td>Abstract</td>
<td>23</td>
</tr>
<tr>
<td>Related publications</td>
<td>23</td>
</tr>
</tbody>
</table>
Executive Summary

This paper describes research conducted for the Department for Transport (Integrated Transport, Economics and Appraisal Division) into the estimation of the number of vehicles that travel from an origin point to a destination point (the O/D throughput) using licence plate data collected using Automatic Number-Plate Recognition (ANPR) systems. The research objectives are:

- To recommend a method of estimating O/D throughput and its accuracy (confidence limits).
- To recommend good practice for the implementation of the recommended method as a free standing document (see Appendix A).

Manually read licence plates have historically been used in surveys to calculate the O/D throughput but in recent years the use of ANPR equipment has risen because large numbers of plates can be read at relatively low cost. ANPR systems generally consist of camera(s) mounted on an over-bridge, gantry or roadside pole facing the on-coming traffic, and plate recognition equipment housed in a roadside cabinet, which sometimes also contains communications equipment.

When manually read plates are used to estimate O/D throughput a simple equation may be used:

\[
OD(AB) = \frac{M(AB)}{R(A)R(B)} \tag{i}
\]

where,

- \(OD(AB)\) is the O/D throughput between sites A and B;
- \(M(AB)\) is the number of all-digit matches (or all-digit plus fuzzy matches\(^1\)) between licence plates read at sites A and B; and
- \(R(A)\) and \(R(B)\) are the proportion of plates read accurately at sites A and B\(^2\).

This equation is satisfactory when unbiased samples of plates are manually read to a high degree of accuracy at different sites and verified by different enumerators. If all-digit plus one-digit fuzzy matching is used with manually read plates it is expected that \(OD(AB)\) can be estimated to a very high degree of accuracy. However, when ANPR systems are used to read the plates a more complex equation is required because ANPR systems do not read 100% of licence plates accurately, nor are the errors confined to one misread character. Plates are completely missed or misread by the ANPR system for reasons that include the character recognition system capability, the vehicle characteristics, including the position and features of the licence plate itself, and the weather conditions.

For a particular vehicle, many of these factors are the same at both sites (for example the system software, vehicle characteristics) and hence it is possibly more likely that the plates of vehicles read accurately at the first site will be read accurately at the second site.

An alternative equation has been derived by TRL to estimate the O/D throughput using ANPR systems:

\[
OD(AB) = \frac{M(AB)}{R(A)R(B)Acc(B)Acc(B)F(B | A)} \tag{ii}
\]

where,

- \(OD(AB)\) is the O/D throughput between sites A and B;
- \(M(AB)\) is the number of all-digit matches between licence plates read at sites A and B;
- \(R(A), R(B)\) are the proportion of plates read at sites A and B;
- \(Acc(A), Acc(B)\) are the proportion of plates read that are accurate at sites A and B; and
- \(F(B | A)\) is an estimate of the ratio \(P(B | A)/P(B)\); where;
- \(P(B)\) is the proportion of plates read accurately at site B; and
- \(P(B | A)\) is the proportion of plates read accurately at site B given that they were read accurately at site A.

The product \(Acc(A)Acc(B)F(B | A)\) is termed the correction factor \(C(AB)\) and its estimation forms the major part of the research. It was expected that \(C(AB)\) would depend on a set of factors (e.g. ANPR manufacturer, % of HGVs, traffic speed, traffic flow etc.). The relationship between \(C(AB)\) and these factors has been investigated using a reference set of data obtained from a programme of video recordings at sites where two manufacturer’s ANPR equipment were installed, so that a correction factor model could be developed. An analytical approach has also been used, based on statistical theory, to establish theoretical relationships between the terms in Equation (ii) and the statistical probability that the plates will be read accurately by the ANPR system. It was intended that the statistical analysis would provide useful insight into the relationship that would help inform the experimental approach. The statistical theory is given in Appendix C of the report.

To develop the relationship, a programme of measurements has been undertaken at six sites located on the M6 and M54 road system in the West Midlands; three with manufacturer X’s equipment and three with manufacturer Y’s. The measurements have included data obtained using video recordings to allow the manual identification of plates. These data were then compared with the ANPR system derived data. The first stage in the process was to manually extract the licence plates from the video recordings and develop a verification method. This

\(^1\) One character different.
\(^2\) If all plates are read then \(R(A)\) and \(R(B)\) are both equal to 1 and \(OD(AB)=M(AB)\).
was implemented for the sites with manufacturer Y’s equipment, since it was found that the number and accuracy of the plates read by manufacturer X’s equipment were inadequate for the purpose of estimating models of OID throughput. The terms in Equation (ii) were then calculated and the relationship with the determining factors studied. It was found that there was little evidence of systematic variation, beyond the site specific features captured in each reference set of licence plate data. The required sample size for the calibration of Equation (ii) was investigated and it was concluded that a large sample of plates (about 1000 observations) is required to determine the correction factor and estimate the OID throughput with an estimated accuracy (calculated from the experimental data) of between ±6% and ±16% at the 95% level. If a higher accuracy is required, a larger sample of plates must be read.

The best practice for the implementation of the method is detailed in Appendix A and includes advice on the equipment specification, collection of the reference set of data and conduct of the OID survey. To use ANPR data for OID throughput estimation, it is necessary to set a high specification for the accuracy of the equipment and to record the licence plate data under conditions that minimise the controllable factors that affect the ability of the system to read the plates accurately, including weather conditions. It is also necessary to have an accurate source of traffic count data available for the sites, as this is needed to calculate the proportion of plates read by the equipment.

The report’s conclusions are as follows:

1. It was found that neither the statistical formulation nor the factor analysis provided a more useful formulation for OID throughput than that provided in Equation (ii), derived in Appendix B. Each system installation requires local ‘calibration’ before use; using a reference set of reliable manually determined data to calculate the terms in the equation, including a correction factor.

2. Based on the findings of this study a method of implementation and validation has been recommended. It is recommended that the method should be used on future installations where ANPR is being used for OID surveys.

3. It would be prudent to consider refining the method in the future, potentially reducing the errors in the correction factor using fuzzy matching (one character different).
1 Introduction

1.1 Objectives
This paper describes research conducted for the Department for Transport (Integrated Transport, Economics and Appraisal Division) into the estimation of the number of vehicles that travel from an origin point to a destination point (the O/D throughput) using licence plate data collected using Automatic Number-Plate Recognition (ANPR) systems. The research objectives are:

- To recommend a method of estimating O/D throughput and its accuracy (confidence limits).
- To recommend good practice for the implementation of the recommended method as a free standing document (see Appendix A).

1.2 Background
Manually read licence plates have historically been used to calculate the O/D throughput but in recent years the use of ANPR equipment has risen because large numbers of plates can be read at relatively low cost. ANPR systems generally consist of camera(s) mounted on an over-bridge, gantry or roadside pole facing the on-coming traffic (see Figure 1.1) and plate recognition equipment housed in a roadside cabinet, which sometimes also contains communications equipment. These systems can read large numbers of licence plates automatically.

Figure 1.1 ANPR installation showing the cameras

When manually read plates are used to estimate O/D throughput a simple equation may be used:

\[ OD(AB) = \frac{M(AB)}{R(A)R(B)} \]  

where:

- \( OD(AB) \) is the O/D throughput between sites A and B;
- \( M(AB) \) is the number of all-digit matches or fuzzy matches between licence plates read at sites A and B; and
- \( R(A) \) and \( R(B) \) are the proportion of plates read at sites A and B.

This equation is satisfactory when unbiased samples of plates are manually read to a high degree of accuracy at different sites and verified by different enumerators. If manually read plates are fuzzy matched it is expected that \( OD(AB) \) can be estimated to a very high degree of accuracy. However, when ANPR systems are used to read the plates a more complex equation is required because ANPR systems do not read 100% of plates accurately, nor are the errors confined to one misread character. Plates are completely missed or misread by the ANPR system for reasons that include the character recognition system capability, the vehicle characteristics, including the position and features of the licence plate itself, and the weather conditions.

For a particular vehicle, many of these factors are the same at the two sites (for example the system software, vehicle characteristics) and hence it is possibly more likely that the plates of vehicles read accurately at the first site will be read accurately at the second site.

An alternative equation has been derived by TRL to estimate the O/D throughput using ANPR data (see Appendix B for derivation):

\[ OD(AB) = \frac{M(AB)}{P(A)P(B | A)} \]  

where:

- \( M(AB) \) is the number of all-digit matches between licence plates read at sites A and B;
- \( P(A) \) is the proportion of plates read accurately at site A; and
- \( P(B | A) \) is the proportion of plates read accurately at site B given that they were read accurately at site A.

An alternative formulation of this equation (also derived in Appendix B), which separates the terms to be calculated and estimated is as follows:

\[ OD(AB) = \frac{M(AB)}{R(A)R(B)Acc(A)Acc(B)F(B | A)} \]  

where:

- \( R(A), R(B) \) are the proportion of plates read at sites A and B;
- \( Acc(A), Acc(B) \) are the proportion of plates read that are accurate at sites A and B; and
- \( F(B | A) \) is an estimate of the ratio \( P(B | A)/P(B) \).

The product \( Acc(A)Acc(B)F(B | A) \) is termed \( C(AB) \), the correction factor.

1 Photos by courtesy of the equipment manufacturers, Golden River Traffic, (www.goldenriver.com).
2 One character different.
3 If all plates are read then \( R(A) \) and \( R(B) \) are both equal to 1 and \( OD(AB)=M(AB) \).
The number of all-digit plus fuzzy matches cannot be substituted for the number of matches in these equations. If \( M(AB) \) is defined as the number of fuzzy matches, the equation is more complex. Fuzzy matching has not been studied in detail in this report, but some preliminary work is included in Appendix D.

### 1.3 Research approach

This paper describes the results of research into the development of a method of estimating the terms in Equation (3), the confidence intervals on the extra terms and the best practice for its implementation. The main approach has been to compare data taken from a reference set of manually extracted plates, obtained from a programme of video recordings, with data obtained at the same location using ANPR systems. Additionally, the study has investigated the relationships between the terms in the equation and parameters that might affect the accuracy of automatically reading number plates. An analytical approach has also been used, based on statistical theory, to establish theoretical relationships between the terms in Equation (3) and the statistical probability that the plates will be read accurately by the ANPR system. It was intended that the statistical analysis would provide useful insight into the relationship that would help inform the experimental approach. The statistical theory is given in Appendix C.

The programme of video recordings took place in January and February 2002 at six locations on the 3-lane M6 and 2-lane M54 in the West Midlands, where two different ANPR manufacturers’ equipment (X and Y) was installed; each at three sites. The six sites at which ANPR systems were already installed are shown in Figure 1.2 and listed as follows:

1. M6 J4 westbound ‘A’ carriageway (manufacturer X);
2. M6 J4A/5 westbound ‘A’ carriageway (manufacturer Y);
3. M6 J7 westbound ‘A’ carriageway (manufacturer Y);
4. M6 J10A northbound ‘A’ carriageway (manufacturer X);
5. M6 J12/13 northbound ‘A’ carriageway (manufacturer Y);

A total of 36,000 plates were extracted manually from the video tapes, with plates from each of 17 separately filmed lanes being extracted in four time periods: morning peak (07:00 to 10:00), daytime off-peak (10:00 to 16:00), evening peak (16:00 to dusk) and night time (hours of darkness). It was recognised that the plates for the same time periods should be extracted for each lane at any one site. An average of approximately 1500 plates were extracted per time period from three lane sites and 1000 plates per time period from the two lane site. Inevitably, this meant that a different number of plates were read from each lane because the number of vehicles using each lane in a fixed time period differs. The video recordings covered a longer period so that further samples could be extracted if necessary. More details of the underlying philosophy for the video recording and plate extraction are given in Section 2.

Information on the parameters that might affect the accuracy of automatically reading the plates was either recorded at the same time as the video recording (e.g. weather conditions) or extracted afterwards from the video tapes (e.g. vehicle types) or traffic monitoring equipment (e.g. vehicle speeds). The system manufacturer was also expected to affect the accuracy because the hardware and software used were different.

### 1.4 Report structure

Section 2 details the initial study, including the video recording trials, the extraction of the licence plates and the initial analysis. A recommended method for extracting licence plates to provide a reference set was developed and implemented as a result of this work. Section 3 of the report explores the relationship between the terms in Equation (3) and their determining factors. A recommended equation to use to estimate the \( O/D \) throughput was calibrated and validated as described in Section 4. Sections 5 and 6 contain the report’s Discussion and Conclusion sections. The recommended method and its implementation were incorporated in the free standing document reproduced in Appendix A. Appendix B contains details of the derivation

### Table 1.1 Origin / destination pairs

<table>
<thead>
<tr>
<th>Destinations</th>
<th>M6 J4 (Site 1)</th>
<th>M6 J4A/5 (Site 2)</th>
<th>M6 J7 (Site 3)</th>
<th>M6 J10A (Site 4)</th>
<th>M6 J12/13 (Site 5)</th>
<th>M54 (Site 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origins</td>
<td>X</td>
<td>Y</td>
<td>X</td>
<td>X</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

Figure 1.2 Diagram of ANPR site locations
of Equations (2) and (3). Appendix C contains the statistical theory and Appendix D contains some initial research into the use of fuzzy matching.

2 Initial study

2.1 Video recording

The philosophy of the TRL video recording trials was to film all the vehicles on the road during daylight hours for a complete day, and half hour samples of vehicles during the hours of darkness. The requirement for the daytime filming was that the licence plate of each vehicle would be clearly visible in the image. Accordingly, each lane was filmed separately. The field of view was set wide enough to include vehicles changing lanes. The overlap was about one metre. The requirement for the night filming of individual lanes was not as stringent, as it was recognised that all vehicles would not be filmed. The field of view was set to record vehicles centred in the lane to maximise the image of the licence plate.

The video recording took place in January and early February 2002 at the sites shown on the map in Figure 1.2. The filming took place in daylight hours (approximately 07:30 to 17:00) and for separate 30 minute periods in each lane at night. The sites filmed simultaneously were:

- Sites 1 and 3: 29th January 2002.
- Sites 2 and 5: 30th January 2002.
- Sites 4 and 6: 31st January 2002 (daylight only) and 5th February 2002 (night only).

Generally three daytime tapes were recorded at each site, commencing at 07:30, with the tapes being changed at about 11:30 and 15:30. Of necessity, the camera was initially set-up for the conditions prevailing at 07:30, when traffic speeds were generally lower than free flow and lighting conditions were poor. A slow shutter speed was set to suit the light conditions. The camera settings were adjusted at 11:30 and 15:30 to suit the prevailing conditions. During the period following 15:30, the light deteriorated and at around 17:00 the filming was halted.

The instructions to the TRL cameramen included a requirement to set the clocks on the video recorder to the accurate time using their ‘Rugby clock’. They were also asked to compile a log of the weather and traffic speed information.

2.2 Licence plate extraction from video

The plan for the plate extraction was to extract about 1500 plates from three lanes (1000 from two lanes on the M54) during the morning peak (07:00 to 10:00), daytime off-peak (10:00 to 16:00) and evening peak periods (16:00 to dusk), as well as 30 minutes of plates at night (during the hours of darkness) from each lane. Each of these periods was recorded on separate videotapes. The periods chosen to extract plates were selected so that the start time at the destination sites (3, 5 and 6) of the pairs filmed simultaneously were off-set by a suitable period from the start times at the origin sites (1, 2 and 4). This ensured that the vehicles passing the origin sites at the start of the extraction period would have passed the destination site during their extraction period, the intention being to match the manually extracted plates and calculate the O/D throughput in order to verify the recommended method.

To decide on the period to be extracted in the morning peak, the start of each of the three morning videos for each of the three lanes at the origin sites were examined and a start time determined. The criterion used was that the licence plate was clearly visible on each video. The traffic flow in one minute in each lane was counted and the length of the counting period determined to ensure that about 1500 plates were extracted, although the length of the period was always a multiple of 5 minutes, because the intention was to divide the plates into 5-minute sets. The same process was applied to the off-peak and evening peak periods, although in the evening, a period of suitable length between 16:00 and 16:30 was chosen, to ensure that the extraction periods at the destination sites were in daylight. The offset for the extraction periods at the destination sites were as follows:

- Site 3: 10 minutes after Site 1.
- Site 5: 20 minutes after Site 2.
- Site 6: 5 minutes after Site 4.

In some cases the plates on the morning peak, evening peak, or occasionally the off-peak videos were not clearly visible and it was decided not to extract plates from the tape.

The plate extraction was conducted at TRL. The instructions to the enumerators was to record the registration plate, time (hours and minutes) and class of vehicle for each plate visible on the tape in the period defined. The class of vehicle was defined as follows:

- 1 Cars: Car based vans, transits and small goods vehicles without ‘cab’ style front.
- 3 Other vehicles: Coaches, cranes etc.

If the characters on the plate were not clear, the enumerators were instructed to replace the unclear character(s) by a question mark.

2.3 Initial analysis of manual and ANPR data

Following the initial extraction of the licence plates and before any validation or correction was implemented, the manually read and automatically read licence plates were compared and the read rate \( R(A) \) and accuracy rate of the plates read \( \text{Acc}(A) \) calculated. Figure 2.1 and 2.2 show the rates for each lane and five minute period.

Both the read rates and the proportion of plates read that are accurate were higher and less variable for the manufacturer Y’s sites. For this manufacturer both \( \text{Acc}(A) \) and \( R(A) \) for sites 5 and 6 were lower at night than during the day, while the reverse was true for the manufacturer X’s sites.

The effect that the varying read and accuracy rate have on the correction factor \( C(AB) \) was investigated using 5-minute sets of plates at each origin and destination using the method detailed in Appendix B. Table 2.1 shows the
average of $C(AB)$, the standard deviation (StDev) and coefficient of variation for 5-minute samples of licence plates, divided into manufacturer at origin (A) and destination (B) and into samples with at least 10 matches between the ANPR licence plates and samples with at least 20 matches between the ANPR licence plates and samples with at least 20 matches.

The coefficient of variation was least for $C(AB)$ based on at least twenty vehicle 5-minute samples, for manufacturer Y’s equipment at both origin and destination. It was greatest for manufacturer X’s equipment at both origin and destination.

As a result of these initial analyses, it was decided that the number and accuracy of the plates read by manufacturer X’s equipment were inadequate for the purpose of estimating models of O/D throughput and that the further development would use data from manufacturer Y’s equipment read in daylight.

\footnote{\textit{R(A) will exceed 1 if there were more plates recorded by the ANPR system than recorded on the video}}
2.4 Licence plate validation

A recommended method for manually extracting and validating licence plates from video film was developed and implemented for the manual plates read at manufacturer Y’s sites. The recommendations are applicable in a wide range of circumstances and are as follows:

1. The extraction of the plates from the video should utilise an enumerator with experience of reading licence plates. If characters are not clear, they should be replaced by a question mark.

2. The sets of manually extracted plates should be matched with the plates read automatically by the ANPR system and two lists of plates and times generated; one for the manually extracted plates and one for the ANPR plates. The plates not matched between the two lists should be highlighted. It can be assumed that the plates that match are accurate.

3. The differences between the manually extracted plates and ANPR plates should be resolved using the two lists, by re-examining the video, and a definitive list of manually extracted plates prepared for each set. If the differences cannot be resolved, the doubtful characters should be replaced by question marks in the plate list. If lighting and other conditions are such that more than 5% of plates are unresolved then the set should not be used.

4. The definitive list should be fuzzy matched (allowing one character difference) against ANPR plates to validate step 3. The number of fuzzy matches would be expected to be lower than for the original two lists.

2.5 Results of correction and validation

The results of the correction and validation process for manufacturer Y’s sites 2, 5 and 6 are shown in Table 2.2, Table 2.3 and Figure 2.3. The tables and figure show the proportion of the manual plates read by the ANPR system in the first and second manual extractions, \((R(A))\), and the proportion that were accurate \((Acc(A))\). Those reported accurate are based on requiring all characters to match when comparing the ANPR and manually extracted records. As expected, the measured accuracy of the ANPR plates was higher after the second manual read, although highly variable, ranging from 0.68 to 0.91 for the first read and from 0.75 to 0.94 for the second read. Figure 2.3 and Table 2.3 also show the proportion of fuzzy matches (one character different), which dropped to below 0.09, except for three sets of data collected at site 5. Table 2.3 shows that the proportion of question marks was also high for these three exceptional sets (above the 5% required) and they have been excluded from the analysis presented in Section 3.

Question marks do not appear in Table 2.2 and Table 2.3 as accurate matches, but where only one digit is replaced by a question mark, that entry is quite likely to lead to what are shown as fuzzy matches in those tables. The variable \((Acc(A) + Fuzzy(A))\) (the proportion of all-digit and fuzzy accuracy) therefore allows for occasions where one digit was replaced by a question mark and seems to be a significant variable in Figure 2.4. The intercept in that figure implies that the maximum accuracy of video extraction for use with one digit fuzzy matching is about 0.97, implying that 3% of the errors in the ANPR plates may not be a single character error and that a third manual read is unlikely to lead to major improvements in accuracy.

Although more research could have been undertaken on fuzzy matching, it is clear, from the analysis reported here, that fuzzy matching represents only a partial solution, which helps indicate the likely improvement from validating manually extracted records.

### Table 2.1 \(C(AB)\) for 5-minute samples for each manufacturer (first extraction)

<table>
<thead>
<tr>
<th>Manufacturer at site A</th>
<th>Daylight or night</th>
<th>Lane at site A</th>
<th>Statistic</th>
<th>(&gt;=10) ANPR matches</th>
<th>(&gt;=20) ANPR matches</th>
<th>(&gt;=10) ANPR matches</th>
<th>(&gt;=20) ANPR matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Daylight</td>
<td>All lanes</td>
<td>Average of (C(AB))</td>
<td>0.87</td>
<td>0.50</td>
<td>0.78</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Std. dev. of (C(AB))</td>
<td>1.11</td>
<td>0.23</td>
<td>0.25</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Samples of (C(AB))</td>
<td>25</td>
<td>2</td>
<td>83</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Coeff. of var. of (C(AB))</td>
<td>128%</td>
<td>46%</td>
<td>32%</td>
<td>38%</td>
</tr>
<tr>
<td>Night</td>
<td>All lanes</td>
<td>Average of (C(AB))</td>
<td>0.93</td>
<td>1.27</td>
<td>0.87</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Std. dev. of (C(AB))</td>
<td>0.42</td>
<td>0.03</td>
<td>0.29</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Samples of (C(AB))</td>
<td>12</td>
<td>2</td>
<td>43</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coeff. of var. of (C(AB))</td>
<td>45%</td>
<td>2%</td>
<td>33%</td>
<td>30%</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>Daylight</td>
<td>All lanes</td>
<td>Average of (C(AB))</td>
<td>0.27</td>
<td>0.27</td>
<td>0.88</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Std. dev. of (C(AB))</td>
<td>0.13</td>
<td>0.13</td>
<td>0.18</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Samples of (C(AB))</td>
<td>155</td>
<td>52</td>
<td>70</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coeff. of var. of (C(AB))</td>
<td>47%</td>
<td>49%</td>
<td>20%</td>
<td>18%</td>
<td></td>
</tr>
<tr>
<td>Night</td>
<td>All lanes</td>
<td>Average of (C(AB))</td>
<td>0.59</td>
<td>0.57</td>
<td>0.94</td>
<td>1.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Std. dev. of (C(AB))</td>
<td>0.13</td>
<td>0.08</td>
<td>0.24</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Samples of (C(AB))</td>
<td>21</td>
<td>6</td>
<td>11</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coeff. of var. of (C(AB))</td>
<td>22%</td>
<td>14%</td>
<td>25%</td>
<td>2%</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.2 Comparisons between manual and ANPR plates before and after correction for manufacturer Y’s sites

<table>
<thead>
<tr>
<th>Site No</th>
<th>Lane</th>
<th>Period</th>
<th>Number of ANPR plates (1st read)</th>
<th>Number of manual plates (1st read)</th>
<th>Number of plates with question marks (1st read)</th>
<th>Number of matches between manual and ANPR plates (1st read)</th>
<th>Number of matches between manual and ANPR plates (2nd read)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>ampeak</td>
<td>306</td>
<td>335</td>
<td>3</td>
<td>262</td>
<td>274</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>inter</td>
<td>355</td>
<td>397</td>
<td>10</td>
<td>5</td>
<td>299</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>pmpeak</td>
<td>228</td>
<td>266</td>
<td>265</td>
<td>9</td>
<td>192</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>ampeak</td>
<td>419</td>
<td>478</td>
<td>482</td>
<td>1</td>
<td>345</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>inter</td>
<td>375</td>
<td>387</td>
<td>394</td>
<td>1</td>
<td>268</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>pmpeak</td>
<td>344</td>
<td>359</td>
<td>360</td>
<td>3</td>
<td>242</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>ampeak</td>
<td>203</td>
<td>244</td>
<td>244</td>
<td>3</td>
<td>175</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>inter</td>
<td>200</td>
<td>252</td>
<td>251</td>
<td>6</td>
<td>178</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>pmpeak</td>
<td>231</td>
<td>272</td>
<td>273</td>
<td>7</td>
<td>163</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>ampeak</td>
<td>353</td>
<td>443</td>
<td>443</td>
<td>11</td>
<td>269</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>inter</td>
<td>319</td>
<td>377</td>
<td>377</td>
<td>19</td>
<td>239</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>pmpeak</td>
<td>394</td>
<td>460</td>
<td>464</td>
<td>17</td>
<td>289</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>ampeak</td>
<td>296</td>
<td>332</td>
<td>332</td>
<td>2</td>
<td>210</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>inter</td>
<td>306</td>
<td>348</td>
<td>347</td>
<td>6</td>
<td>250</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>pmpeak</td>
<td>455</td>
<td>521</td>
<td>520</td>
<td>4</td>
<td>405</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>inter</td>
<td>468</td>
<td>511</td>
<td>514</td>
<td>5</td>
<td>405</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>inter2</td>
<td>530</td>
<td>573</td>
<td>577</td>
<td>3</td>
<td>450</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>pmpeak</td>
<td>476</td>
<td>591</td>
<td>597</td>
<td>7</td>
<td>363</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>inter</td>
<td>323</td>
<td>367</td>
<td>368</td>
<td>5</td>
<td>248</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>inter2</td>
<td>396</td>
<td>456</td>
<td>433</td>
<td>0</td>
<td>351</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>pmpeak</td>
<td>436</td>
<td>656</td>
<td>647</td>
<td>10</td>
<td>322</td>
</tr>
</tbody>
</table>

Table 2.3 Accuracy and other proportions before and after correction for manufacturer Y’s sites

<table>
<thead>
<tr>
<th>Site No</th>
<th>Lane</th>
<th>Period</th>
<th>Acc(A) (2nd read)</th>
<th>Acc(A) (1st read)</th>
<th>R(A) (2nd read)</th>
<th>R(A) (1st read)</th>
<th>Improvement in Prop of fuzzy matches (1st to 2nd read)</th>
<th>Proportion of question marks (2nd read)</th>
<th>Proportion of question marks (1st read)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>ampeak</td>
<td>0.9</td>
<td>0.86</td>
<td>0.91</td>
<td>0.91</td>
<td>0.11</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>inter</td>
<td>0.87</td>
<td>0.86</td>
<td>0.89</td>
<td>0.89</td>
<td>0.11</td>
<td>0.98</td>
<td>0.96</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>pmpeak</td>
<td>0.94</td>
<td>0.87</td>
<td>0.86</td>
<td>0.86</td>
<td>0.11</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>ampeak</td>
<td>0.89</td>
<td>0.83</td>
<td>0.88</td>
<td>0.88</td>
<td>0.11</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>ampeak</td>
<td>0.89</td>
<td>0.89</td>
<td>0.9</td>
<td>0.9</td>
<td>0.08</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>inter</td>
<td>0.93</td>
<td>0.89</td>
<td>0.85</td>
<td>0.85</td>
<td>0.08</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>pmpeak</td>
<td>0.89</td>
<td>0.89</td>
<td>0.85</td>
<td>0.85</td>
<td>0.08</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>ampeak</td>
<td>0.93</td>
<td>0.75</td>
<td>0.95</td>
<td>0.95</td>
<td>0.05</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>ampeak</td>
<td>0.87</td>
<td>0.78</td>
<td>0.96</td>
<td>0.96</td>
<td>0.09</td>
<td>0.97</td>
<td>0.95</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>ampeak</td>
<td>0.94</td>
<td>0.87</td>
<td>0.83</td>
<td>0.83</td>
<td>0.01</td>
<td>0.95</td>
<td>0.93</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>inter</td>
<td>0.88</td>
<td>0.78</td>
<td>0.96</td>
<td>0.96</td>
<td>0.17</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>ampeak</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.05</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>Inter2</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.07</td>
<td>0.95</td>
<td>0.93</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>pmpeak</td>
<td>0.93</td>
<td>0.89</td>
<td>0.91</td>
<td>0.91</td>
<td>0.03</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>Inter2</td>
<td>0.89</td>
<td>0.89</td>
<td>0.91</td>
<td>0.91</td>
<td>0.04</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>pmpeak</td>
<td>0.94</td>
<td>0.75</td>
<td>0.67</td>
<td>0.66</td>
<td>0.03</td>
<td>0.97</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Average: 0.89 | 0.82 | 0.87 | 0.87 | 0.06 | 0.11 | 0.96 | 0.93 | 0.02 | 0.02 | 0.02
Average excluding highlighted values: 0.91 | 0.83 | 0.87 | 0.87 | 0.05 | 0.10 | 0.96 | 0.93 | 0.02 | 0.01 | 0.02
Figure 2.3 $R(A)$, $Acc(A)$ and proportion of fuzzy matches between manual and automatic plates after first and second read of manual plates.

Figure 2.4 Improvement between first and second manual reads of video recorded number plates.
3 Relationship with determining factors

3.1 Determining factors

The accuracy of automatically reading number plates could depend on many factors which may include:

- The type of equipment used.
- Traffic composition.
- Speed.
- Flow density/lane occupancy.
- Weather conditions.
- Time of day.
- Site layout.

In order to establish the importance of these parameters, data was collected and analysed from the three sites with manufacturer Y’s equipment and the validated manually extracted plates, after the second extraction of plates, were used to provide a reference set. Traffic data at each site was collected from Highways Agency traffic counting inductive loops and weather data was collected from the Met. Office Coleshill weather station.

3.2 Methodology

The relationship between the determining factors and the terms $\text{Acc}(A)$, $\text{Acc}(B)$, $F(B \mid A)$ in Equation (3), and their product $C(AB)$, the correction factor, was investigated. The manually extracted plates at each site were used to estimate $\text{Acc}(A)$ and $\text{Acc}(B)$ by direct comparison between the plates read by the ANPR system and the corresponding ‘second read’ version of the manually read set. $M(AB)$ was calculated by matching the plates read by the ANPR systems. $F(B \mid A)$ was estimated using the following equation (see Appendix B for derivation):

$$F(B \mid A) = \frac{X}{P(A)}$$

where,

- $X$ is the ratio $M(AB)/MD(AB)$;
- $M(AB)$ is the number of matches between the ANPR read plates at site A and the ANPR system plates at site B;
- $MD(AB)$ is the number of matches between the manually read plates at site A and the ANPR system plates at site B; and
- $P(A)$ is the proportion of plates read accurately at site A.

3.3 Results of analysis

Figure 3.1 shows the values of $\text{Acc}(A)$ for 5-minute samples of plates read by manufacturer Y’s equipment using validated manually extracted licence plates after the second extraction (see Section 2). $\text{Acc}(A)$ is shown to vary within the time periods and within lane group. There is no evidence of a relationship between $\text{Acc}(A)$ and either time of day or lane during daylight hours.

Figures 3.2 and 3.3 show examples of the relationships in daylight between the terms $\text{Acc}(A)$ and $C(AB)$ in Equation (3) and their determining factors. The effect of weather was not studied because the survey days were generally dry and clear. Figure 3.2 shows that there is little evidence of a relationship between $\text{Acc}(A)$ and the postulated determining factors. Very little of the variation in $\text{Acc}(A)$ is explained by these factors, although there is a very slight tendency for $\text{Acc}(A)$ to increase with average

![Figure 3.1](image.png)
Figure 3.2 Plots of relationship between accuracy rate $\text{Acc}(A)$ and traffic flow, speed, occupancy$^5$ and proportion of HGVs (5-minute samples in individual lanes for manufacturer Y)

Figure 3.3 Plots of relationship between correction factor $(C(AB))$ and traffic flow, speed, occupancy and proportion of HGVs (5-minute samples in individual lanes for manufacturer Y)

speed. However, the trend line is heavily affected by a single high value. Figure 3.3 shows a similar pattern for $C(AB)$ with very little evidence of a relationship with any of the factors studied, although there is again a tendency for $C(AB)$ to increase with average speed.

$^5$ Occupancy is the percentage of a minute period that vehicles occupy the traffic counting inductive loop
4 Model calibration and validation

4.1 Proposed model

As a result of the investigation in Section 3, it was evident that a model of O/D throughput could not be based on a relationship with the potential determining factors identified, and a simple model was proposed:

\[ OD(AB) = \frac{M(AB)}{R(A)R(B)C_{REF}(AB)} \]  

(4)

where \( C_{REF}(AB) \) is based on an estimate derived from a reference set of data:

\[ C_{REF}(AB) = F(B|A)Acc(A)Acc(B) \]  

(5)

The following sections report the investigation of the required sample size for the reference set for the model calibration and reports on the validation of the recommended model.

4.2 Model calibration

The calculation of \( C_{REF}(AB) \) requires a manual extraction of licence plates from a video and an ANPR system for the same period at either site A or site B. Figure 4.1 shows values of \( C_{REF}(AB) \) (labelled \( C(AB) \)) and its components in Equation (5) for 5-minute samples of licence plates read by manufacturer Y’s equipment. Licence plates are read at the A sites and matched with B sites, with the B sites defined as the second match - which can be either before or after A, doubling the sample available to calibrate the model.

The plot shows that while \( Acc(A) \) and \( Acc(B) \) were generally stable, values of \( F(B|A) \) and hence the product \( C_{REF}(AB) \) were not. The mean value of \( Acc(A) \) (and, since all sites were both A and B sites, \( Acc(B) \)) was 0.91 and the standard deviation 0.03. The mean value of \( C_{REF}(AB) \) was 0.93 and the standard deviation was 0.12.

Figure 4.2 and Table 4.1 shows the values of \( C_{REF}(AB) \) (labelled \( C(AB) \)) and its components for the whole of each typically 20-minute period, representing the morning peak, inter-peak or afternoon peak periods. The figures are based on larger samples and the standard deviation of \( C_{REF}(AB) \) is consequently lower: 0.06. The mean value is 0.95.

The values of \( C_{REF}(AB) \) varied from 0.84 to 1.05. An assumption that \( C_{REF}(AB) \) is unity could result in an error in the calculation of the O/D throughput of up to 16%. The reference sets for sites 2 and 5 were collected on the same day, while the set for site 6 was collected on a different day. Hence, it is not unexpected that there is a greater similarity between the estimates of \( C_{REF}(AB) \) for Site 2 to Site 5 and Site 5 from Site 2 than between the other pair of sites, since the parameters were more likely to have been estimated from the same vehicle sample.

Since \( C_{REF}(AB) \) varied between pairs of sites it is recommended that each pair of sites is individually calibrated and the calibration is based on a large number of plates, of the order of 1000. The coefficient of variation of \( C_{REF}(AB) \) for each pair of sites in the three periods ranged from 3% to 8% and thus using a calibration based on one of these periods, for example using the value from the ‘inter’ period as an estimate of \( C_{REF}(AB) \) for the ‘ampeak’ or ‘pmpeak’ period, could result in an error in the O/D throughput, assuming that \( R(A) \) and \( R(B) \) are accurately estimated, of between ±6% and ±16% at the 95% level; assuming a normal distribution.

4.3 Model validation

The model was validated using a reserved set chosen as the 331 vehicles that passed site 2 between 11:35 and 11:40.

![Figure 4.1 Variability of components of C_{REF}(AB) for 5-minute samples in all lanes](image)
The list of manually read plates at site 5 indicated that 35 of these passed site 5 but there were only 25 matches between the plates read by the ANPR systems. The terms in the estimation Equation (4) are as follows:

\[ M(AB) = 25; \]
\[ R(A) = 0.92; \]
\[ R(B) = 0.84; \]
\[ C_{REF}(AB) = 0.96. \]

Thus using Equation (4) gives,

\[ OD(AB) = 33.7. \]

The estimated \( O/D \) is thus 34, rounded to a whole number, and the error in this validation is 3\% (well within the expected confidence limit). The error if \( C_{REF}(AB) \) is assumed to be unity is 8\%.

### 5 Discussion

A method of estimating \( O/D \) throughput using ANPR data has been derived using the results of the research described in this report. It had been expected that the model for \( O/D \) throughput would contain a correction factor term, dependent on the values of a set of parameters measured at the time of the \( O/D \) survey relative to those measured when a reference set of plates were recorded by video filming. However, it was found that there was no significant dependency between the correction factor and the expected factors, such as traffic flow, speed and the proportion of HGVs. Hence, the proposed model was simple, containing a correction factor term, estimated from a reference set of data at origin and destination sites. This reference set would be used to calibrate the model for each \( O/D \) pair.
Confidence limits have been estimated from the experimental data collected and may not be transferable to other sites. The limits were estimated to be between ±6% and ±16% at the 95% level. The actual error found from a validation process was 3%. Without the use of a correction factor, the error found would be higher, as found for the reserved set used for validation.

The recommended method and the good practice for its implementation are detailed in Appendix A. An example of its use is included. To maximise the accuracy of the estimates it is recommended that steps are taken to control the factors likely to affect the accuracy. These include the specification for the ANPR equipment, the location of the cameras and the choice of day and time for both the collection of the reference set and the survey. The estimates will be most accurate if the ANPR equipment is capable of reading at least 80% of plates accurately under daylight conditions, in dry weather and with visibility in excess of 500 metres. The reference set video filming and survey should be conducted under these conditions. The camera sites should be carefully chosen, away from locations where traffic is likely to be queuing at the time of collection of the reference set and survey data.

The survey to collect reference data will involve matching the licence plates between pairs of sites for appropriate origin interval lengths; generally these depend on the end use of the O/D data. The origin intervals will need to be long enough for a reasonable number of vehicles to have travelled to each of the required destinations. This inevitably means that some of the vehicles will not have travelled directly, having made a diversion or stopped for a period on the route. If these outlier vehicles are to be excluded, then the data need to be cleaned by examining the journey times of vehicles and the likely maximum journey time for a vehicle travelling directly at the chosen time of day.

Fuzzy matching (one character different) is a technique often used when manually read plates are used for O/D surveys. The errors in reading plates in manual surveys are more likely to be confined to one character, than is the case with ANPR surveys. The analysis of manual and ANPR plates reported in Section 2 suggests that 3% of errors remain after fuzzy matching. The errors incurred by incorrectly matching plates by fuzzy matching throughout have not been considered. Further research is required into the errors involved in the correction factors if fuzzy matching is used. The initial research detailed in Appendix D suggests that a correction factor will be required.

6 Conclusions

To develop a reliable method of estimating O/D throughput using licence plate data read by Automatic Number Plate Recognition (ANPR) equipment, a programme of data collection has been undertaken at six sites located on the M6 and M54 road system in the West Midlands. The measurements have included data obtained using video recordings to allow the manual identification of plates. This data has then been compared with the data derived from the ANPR system.

The study has also examined the relationships between ANPR accuracy and factors such as vehicle type, traffic flow and speed. A theoretical statistical analysis has been performed to determine the relationship between the number of matches between licence plates read by the ANPR systems at two sites and the O/D throughput.

The research conclusions are as follows:

1. It was found that neither the statistical formulation nor the factor analysis provided a more useful formulation for O/D throughput than that provided in Equation (3), derived in Appendix B. Each system installation requires local ‘calibration’ before use; using a reference set of reliable manually determined data to calculate the terms in the equation, including a correction factor.

2. Based on the findings of this study a method for implementation and validation has been recommended. It is recommended that the method should be used on future installations where ANPR is being used for O/D surveys.

3. It would be prudent to consider refining the method in the future, potentially reducing the errors in the correction factor using fuzzy matching (one character different).

7 Acknowledgements

The work described in this report was carried out in the Transport Systems Unit of TRL Limited. The authors are grateful to David Eves who carried out the quality review and auditing of this report.
Appendix A: Using Automatic Number Plate Recognition (ANPR) equipment to estimate Origin to Destination (O/D) throughput using all-digit matching

Background
Manually read licence plates have historically been used to calculate O/D throughput, whereas ANPR data is generally seen as too biased for that use because of correlation between the proportions of plates read accurately at two sites, but unbiased when used to calculate journey times. Both techniques involve matching licence plates read at different locations on the road network. At sites with bridge or gantry-mounted cameras, these generally face the on-coming traffic and the plates of a proportion of the vehicles that pass underneath the cameras are read accurately. Plates are inevitably completely missed and misread for reasons that include the character recognition system capability, the vehicle characteristics, including the position and features of the licence plate itself, and the weather conditions.

In order to use ANPR data for O/D throughput estimation it is necessary to set a high specification for the accuracy of the equipment and to record the licence plate data under conditions that minimise the controllable factors that affect the ability of the system to read the plates accurately, including weather conditions. It is also necessary to have an accurate source of traffic count data available for the sites, as this is needed to calculate the proportion of plates read by the equipment.

The O/D survey design needs to include the selection of sites and time periods and be accompanied by validation tests and analysis to justify using the equipment. O/D data between those sites can then be obtained and extrapolated by other procedures to represent the desired journey. One of the sites in each element of the analysis is called the A site and second sites are referred to as B sites.

The licence plates of vehicles at the A sites in the selected time period need to be read by the ANPR equipment and matched with the licence plates at the B sites. Because the time taken for vehicles of different types to travel between sites varies and there is a possibility of congestion affecting the journey, the time period window at the B sites needs to be longer than that chosen at the A sites. This unavoidable aspect means that some vehicles do not travel directly between sites are included in the matched data set. These vehicles are termed outliers and, if required, should be removed by inspecting the journey time data

The method of estimating the O/D throughput for each pair of sites A and B, OD(AB) uses the following equation:

\[
OD(AB) = \frac{M(AB)}{R(A)R(B)C(AB)} \tag{a}
\]

The terms in this equation are as follows:

- **M(AB)**: The number of all-digit matches between the licence plates read by the ANPR system at sites A and B, cleaned for outliers if necessary.
- **R(A)**: The proportion of vehicles whose ANPR plates have been read by the equipment at site A during the origin period.
- **R(B)**: The proportion of vehicles whose ANPR plates have been read by the equipment at site B during the period that the matched vehicles passed site B.
- **C(AB)**: A correction factor, required because the plates read accurately at site A are more likely to be read accurately at site B.

The factor C(AB) is expected to be close to unity for high-performance ANPR systems when M(AB) is based on an all-digit match. The terms R(A) and R(B) can be estimated from the independent traffic count data but C(AB), which depends on the characteristics of the ANPR systems and the quality of the installation, needs to be estimated from information collected during independent filming of the traffic at the ANPR sites. That independent filming is required for a reference set of about 1000 vehicles, whose plates need to be extracted manually. That process of manual extraction is likely to introduce errors unless a procedure is followed to validate the manually extracted plates.

There are a number of sources of error in the estimate of OD(AB). These are:

- errors in the traffic count data;
- errors in the manual reading of reference set plates;
- errors in the correction factor C(AB) due to its calculation being based on a limited reference data set.

The magnitude of the first two types of error can be minimised by validating the manually read plates and using a count of plates from the video records to check the accuracy of the main source of traffic count data. The error in the correction factor can be minimised by appropriate design of the O/D survey and validation filming and calculations as discussed below.

Summary of procedure
The stages in the procedure for estimating the O/D throughput from ANPR data and its errors are outlined below and each stage is detailed in the following sections:

A System and Installation: Install ANPR equipment at each site identified in the traffic survey design.

B Reference data set and its validation: Film for 30 minutes (or longer under certain traffic conditions) at each site with each lane at a particular site being filmed simultaneously. Manually extract plates for a fixed time period; a total of about 1000 plates from all lanes. Verify the plates using the recommended method.
C Main O/D survey using ANPR data: Select the ANPR survey data set(s) for which the O/D throughput is required. Match the site A and site B ANPR records. Clean the data for outliers if necessary and calculate the number of matched vehicles $M(AB)$.

D Estimating read rates and correction factor: Extract traffic count data for the periods of the O/D survey and calculate the read rates for each A and B site ($R(A)$ and $R(B)$) for the required time periods. Use the reference set to estimate the accuracy of the traffic count data. Use the reference set to estimate the correction factor $C(AB)$ and its accuracy using the recommended method.

E Estimating OD(AB) and reporting confidence levels: Use the Equation (a) to estimate the O/D throughput for each pair of sites.

A: System and installation
The recommended performance for the ANPR equipment is that at least 80% of plates should be read accurately under daylight conditions, in dry weather and with visibility in excess of 500 metres. This requirement can be tested using the reference data set at each site (see Section B). The system time stamp should be accurate to within 5 seconds of the true value. This is necessary for accurately selecting the corresponding set of traffic count data and determining outliers.

The choice of sites (either over-bridges or gantries) should be made carefully. It is extremely unwise to carry out surveys between points where traffic is stationary at the time that the survey is to be conducted as the plates may be obscured by following vehicles, particularly by Heavy Goods Vehicles. It is desirable that the number of vehicles weaving at the site is small and it is thus recommended that sites are away from a junction (either upstream or downstream) or a location where the number of lanes change, for example for a climbing lane. It is also desirable that if possible the orientation of the cameras is not facing the sun at the time of day that the survey is to be conducted as the shadow cast by the vehicle can cause the ANPR system to trigger on the leading edge of the shadow rather than on the front of the vehicle.

The ANPR cameras should be installed to read all the vehicles on the road. This will generally require one camera located above each lane and the field of view to overlap so that the plates of some weaving traffic can be read.

The camera lens or cover plate should be cleaned immediately prior to the collection of the reference set of data and during the survey itself.

To summarise:
- System 80% accurate in daylight, dry weather and visibility over 500 metres.
- System clock accurate to within 5 seconds.
- Site without stationary traffic, away from a junction or lane change location.
- Cameras installed to read all vehicle plates (overlapping lanes if necessary).
- Lenses cleaned immediately prior to reference filming and surveys.

B: Reference data set and its validation
The reference data set should be extracted from an independent video recording of the licence plates of vehicles in each lane at each of the ANPR sites. The video recording for a particular site should be filmed in each lane simultaneously in daylight conditions, dry weather with visibility greater than 500 metres. If possible the conditions should be bright without any direct sunlight. Times with queuing or slow-moving vehicles should not be used if at all possible because the licence plates of some vehicles are more likely to be obscured by following vehicles under these conditions.

The video cameras should be set to record the licence plates of all vehicles by overlapping the field of view to permit the plates of weaving vehicles to be visible. The timestamp on the video recorder should be set accurately to the nearest five seconds using the TIM time signal or some other means.

The length of filming depends on the traffic flow and should be sufficient to allow about 1000 (between 900 and 1100) licence plates to be read during a fixed time period from all lanes. The time period should be an appropriate fixed number of minutes to permit the comparisons with traffic count data.

The procedure for manually extracting and validating the plates is as follows:

1. An enumerator with experience of reading licence plates should extract the required plates and associated time stamp from the video recording. If characters are not clear, they should be replaced by a question mark.

2. The manually extracted plates should be matched with plates read automatically by the ANPR system at the same site and two lists of plates and times generated; one for the manually extracted plates and one for the ANPR plates. The plates not matched between the two lists should be highlighted. It can be assumed that the plates that match are accurate.

3. The accuracy of the manually extracted plates should be resolved by an independent experienced enumerator, using the two lists, by re-examining the video recording, and a definitive list of manually extracted plates prepared. If more than 5% of the plates are unresolved the reference set is not suitable and the procedure should be repeated.

4. The lists for each lane should be compared and any duplicates removed.

To summarise:
- Record video of each lane simultaneously in daylight (preferably without sun), dry weather and visibility greater than 500 metres without queuing or slow-moving traffic.
- Over-lap the field of view of the cameras.
- Set the timestamp on the video recorder to the nearest second.
- Extract 900 to 1100 plates for a fixed time period in all lanes.
- Validate the manual plates and remove duplicates.
C: Main O/D survey using ANPR data

The time of day and time interval for the Site A data set will be dictated by the survey design. The ANPR data for the whole of the survey day or days should be recorded. The licence plate data for the required periods should be extracted. The licence plate data for a longer period should be extracted for each B site; the length being determined by the maximum acceptable journey time. It is recommended that initially the Site A sets are matched with the Site B sets for the complete day and the maximum acceptable journey time determined by plotting this journey time data. The Site B set can then be limited and the data re-matched and an initial estimate of \( M(AB) \) made. The time periods for which matching are conducted are termed the A and B data set periods. It is assumed that the software will not permit duplicate matches if a licence plate is recorded more than once.

Although this process will reduce the number of outliers, the matched data set will inevitably include some vehicles that did not travel directly between the sites. There are a number of possible methods of cleaning the data for outliers, but it is recommended that if possible the data is cleaned manually. Most automatic methods remove a percentage of data records; assuming that a vehicle whose journey time is more than 2 or 3 standard deviations from the mean journey time for vehicles in the same cohort of a fixed number of vehicles, or in the same 5-minute period, is an outlier. Although for journey time analyses this elimination in both before and after surveys may not affect the accuracy of the results, for \( O/D \) estimation an error may be introduced into the process.

The manual cleaning process depends on the number of matches, the potential for diversion between the sites and the journey pattern. Factors to take into account are:

1. The minimum journey time (perhaps at 80 mph).
2. The presence of a service area on the route and the number of junctions.
3. The similarity in journey time for the majority of vehicles originating in the same lane.

The difficulties of an automatic method are best illustrated with an example. Suppose there are ten matches for vehicles originating in the same 5-minute period on a journey of 20 km passing two junctions. If the journey time for eight of them is 15 to 19 minutes (mean journey time 17 minutes and standard deviation say 1 minute, 95% confidence limits 13 to 21 minutes) and for two of them is 22 minutes, an automatic method would remove the latter pair. However, the longer journeys are unlikely to be outliers as an additional 3 minutes would be insufficient for a vehicle to leave the road and return.

The number of outliers determined should be subtracted from the initial estimate of \( M(AB) \) and the final value obtained.

To summarise:

- Select site A periods.
- Match A data set with all-day B data sets.
- Restrict Site B sets to maximum journey time and re-match to initially estimate \( M(AB) \).
- Manually remove outliers to finalise \( M(AB) \).

D: Estimating read rates and correction factor

For the time period related to each set of about 1000 plates in the reference set, extract the traffic count data. Calculate the ratio of traffic counts to the count of plates in the reference set to estimate the accuracy of the traffic count data. For the A and B data set periods (derived in Section C), extract the traffic count data. The ratio of the number of plates read by the ANPR systems at each site to the corresponding traffic counts gives an estimate of \( R(A) \) and \( R(B) \) for the time periods related to each set. This can be adjusted according to any estimated error in the traffic count data.

The correction factor \( C(AB) \) is estimated by calculating the factor for the reference set. The method requires four quantities to be calculated. For each pair of sites these are:

- \( MD(AB) \) - The number of matches (cleaned using the same method as described in Section C) between the reference set of about 1000 plates read manually from video recordings at site A and the ANPR data at site B.
- \( M(AB) \) - The number of matches (cleaned using the same method as described in Section C) between the ANPR set equivalent to the reference set at site A and the ANPR data at site B.
- \( Acc(B) \) - The proportion of ANPR plates read at site B that are accurate (obtained from the reference set data at site B).
- \( R(A) \) - The proportion of vehicles whose ANPR plates had apparently been read at site A (obtained from the reference set data at site A).

The equation used to calculate \( C(AB) \) is as follows:

\[
C(AB) = \frac{M(AB) \ Acc(B)}{MD(AB) \ R(A)} \quad \text{(b)}
\]

To summarise:

- Calculate the terms \( MD(AB) \), \( M(AB) \), \( Acc(B) \) and \( R(A) \).
- Estimate \( C(AB) \) using Equation (b).
E: Estimating OD(AB) and reporting confidence levels

The terms in the estimation Equation (a) for OD(AB) are estimated by following the above processes:

- \(M(AB)\) as described in Section C;
- \(R(A)\) and \(R(B)\) estimated as described in Section D; and
- \(C(AB)\) estimated using Equation (b) of Section D.

Substitute in Equation (a) to estimate OD(AB):

\[
OD(AB) = \frac{M(AB)}{R(A)R(B)C(AB)}
\]  

(a)

The expected confidence limits for OD(AB) has been estimated from experimental data as between ±6% and ±14%.

An example of the process is given below for three sites. Vehicles can travel from site 1 to site 2 and from site 1 to site 3 but not from site 2 to site 3. The tables show the values used to estimate \(C(AB)\) and \(OD(AB)\).

<table>
<thead>
<tr>
<th>Manual Site A</th>
<th>Site B</th>
<th>Manual count at site A</th>
<th>Manual MD(AB)</th>
<th>R(A)</th>
<th>R(B)</th>
<th>Acc(B)</th>
<th>C(AB)</th>
<th>OD(AB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1236</td>
<td>125</td>
<td>131</td>
<td>0.91</td>
<td>0.82</td>
<td>0.92</td>
<td>0.96</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1360</td>
<td>51</td>
<td>51</td>
<td>0.91</td>
<td>0.70</td>
<td>0.95</td>
<td>1.05</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>847</td>
<td>99</td>
<td>108</td>
<td>0.82</td>
<td>0.91</td>
<td>0.90</td>
<td>1.01</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1403</td>
<td>146</td>
<td>218</td>
<td>0.70</td>
<td>0.91</td>
<td>0.87</td>
<td>0.84</td>
</tr>
</tbody>
</table>
Appendix B: Derivation of equations

If licence plates are read by either an ANPR or manual system, the number of plates matched (all-digits) between the two sites \( M(AB) \) is equal to the number of plates of vehicles passing both sites that are read accurately at site A \( (N(A)) \) multiplied by the proportion of these plates read accurately at site B. This latter quantity is given the symbol \( P(B \mid A) \) and defined as the proportion of plates read accurately at site B, given that they were read accurately at site A. The equation for \( M(AB) \) is thus:

\[
M(AB) = N(A)P(B \mid A) \tag{B1}
\]

The quantity \( N(A) \) is equal to the \( O/D \) throughput \( (OD(AB)) \) multiplied by the proportion of the plates of these through vehicles read accurately at site A \( P(A) \). Substituting for \( N(A) \) in Equation (B1) and rearranging gives the following equation for \( OD(AB) \):

\[
OD(AB) = \frac{M(AB)}{P(A)P(B \mid A)} \tag{B2}
\]

where,

\( P(A) \) is the proportion of plates read accurately at site A; and

\( P(B \mid A) \) is the proportion of plates read accurately at site B given that they were read accurately at site A.

This equation is Equation (2) in Section 1. The terms in the denominator of this equation relate to the through vehicles but can be estimated from data relating to the vehicles passing site A and those passing site B.

An alternative formulation of Equation (B2), which separates the terms to be calculated and estimated, is as follows:

\[
OD(AB) = \frac{M(AB)}{R(A)R(B)Acc(A)Acc(B)F(B \mid A)} \tag{B3}
\]

where,

\( M(AB) \) is the calculated number of matches between plates read at sites A and B;

\( R(A) \) and \( R(B) \) are estimates of the proportion of plates read at sites A and B;

\( Acc(A) \) and \( Acc(B) \) are estimates of the proportion of plates read that are accurate at sites A and B; and

\( F(B \mid A) \) is an estimate of the ratio \( P(B \mid A)/P(B) \).

This equation is Equation (3) in Section 1. The product of the last three terms in the denominator is termed the correction factor \( C(AB) \) and hence

\[
C(AB) = Acc(A)Acc(B)F(B \mid A) \tag{B5}
\]

The quantities \( Acc(A) \) and \( Acc(B) \) can be estimated from a reference set of plates read manually at each site and \( F(B \mid A) \) is estimated by the following process:

1. Match the manually read plates at site A with the ANPR system plates at site B \( (MD(AB)) \), which is equal to \( OD(AB)P(B) \).
2. Match the ANPR system plates at site A with the ANPR system plates at site B \( (M(AB)) \), which is equal to \( OD(AB)P(A)P(B \mid A) \).
3. Since \( P(A) = R(A)Acc(A) \) can be estimated, the following equation can be used to estimate \( F(B \mid A) \):

\[
F(B \mid A) = \frac{X}{P(A)} = \frac{M(AB)}{P(A)MD(AB)} \tag{B4}
\]

The correction factor \( C(AB) \) is given by the following equation:

\[
C(AB) = \frac{Acc(A)Acc(B)M(AB)}{P(A)MD(AB)}
\]

Since \( P(A) = R(A)Acc(A) \) this reduces to

\[
C(AB) = \frac{M(AB) Acc(B)}{MD(AB) R(A)} \tag{B5}
\]

\( \text{Note that } P(A) = R(A)Acc(A) \) and \( P(B) = R(B)Acc(B) \)
Appendix C: Statistical theory

In this appendix it is proved that, under certain plausible conditions, the expected value of \( P(B \mid A) \) in Equation (B2) of Appendix B is greater than \( P(B) \) and the ratio \( F(B \mid A) \) is therefore expected to be greater than unity. The probability that a plate is read accurately at site A is given the symbol \( p_i(A) \) for the ith vehicle. The \( p_i(A) \) are assumed to be independent; that is the probability of a particular plate being read accurately is not affected by any other vehicle. \( F(B \mid A) \) can be estimated from the statistical distribution of these probabilities.

For the \( k \) through vehicles \((k = \text{O/D throughput } OD(AB))\) passing sites A and B, where there is a second set of ANPR cameras, the expected proportions of plates that are correctly read, \( E(P(A)) \) and \( E(P(B)) \), are given by the following equations:

\[
E(P(A)) = \frac{\sum_{i=1}^{k} p_i(A)}{k} \quad \text{(C1)}
\]

\[
E(P(B)) = \frac{\sum_{i=1}^{k} p_i(B)}{k} \quad \text{(C2)}
\]

The expected number of matches between plates read at sites A and B, \( E(M(AB)) \), is given by the following equation:

\[
E(M(AB)) = \sum_{i=1}^{k} p_i(A)p_i(B) \quad \text{(C3)}
\]

It is postulated that the probability of a particular plate being read accurately by a particular ANPR system at sites A and B is different for each plate but equal to fixed site probabilities, \( s(A) \) and \( s(B) \), incorporating various random components, multiplied by an independent individual plate probability \( p_i \), that is:

\[
p_i(A) = s(A)p_i \quad \text{and} \quad p_i(B) = s(B)p_i
\]

The mean value of the \( p_i \) (given the symbol \( \mu \)) is given by:

\[
\mu = \frac{\sum_{i=1}^{k} p_i}{k}
\]

Substituting for \( p_i(A) \) and \( p_i(B) \) in Equations (C1), (C2) and (C3) gives:

\[
E(P(A)) = s(A)\frac{\sum_{i=1}^{k} p_i}{k} = s(A)\mu;
\]

\[
E(P(B)) = s(B)\frac{\sum_{i=1}^{k} p_i}{k} = s(B)\mu
\]

\[
E(M(AB)) = s(A)s(B)\sum_{i=1}^{k} p_i^2; \quad \text{(C4)}
\]

\[
E(P(A))E(P(B))\sum_{i=1}^{k} p_i^2 = \frac{\sigma^2}{\mu^2}
\]

The variance of the \( p_i \) (given the symbol \( \sigma^2 \)) is given by the following equation:

\[
\sigma^2 = E(p_i^2) - \mu^2;
\]

\[
= \frac{\left( \sum_{i=1}^{k} p_i^2 - k\mu^2 \right)}{k}
\]

Substituting in Equation (C4) for \( \sum_{i=1}^{k} p_i^2 \) and replacing \( k \) by \( OD(AB) \) gives the following equation:

\[
E(M(AB)) = OD(AB)E(P(A))E(P(B))\frac{\left( \sigma^2 + \mu^2 \right)}{\mu^2}
\]

Re-arranging Equation (B2) in Appendix B and using expected values gives:

\[
E(M(AB)) = OD(AB)E(P(A))E(P(B \mid A)) \quad \text{(C5)}
\]

as the terms are independent. Hence,

\[
E(P(B \mid A)) = E(P(B))\frac{\left( \sigma^2 + \mu^2 \right)}{\mu^2} \quad \text{(C6)}
\]

This equation implies that under the stated assumptions, \( P(B \mid A) \) is expected to be greater than \( P(B) \). While the expected values of \( P(A) \) and \( P(B) \) depend on the mean values of the accurate read probabilities, the expected value of \( P(B \mid A) \) depends on the variance. Hence the magnitude of \( P(B \mid A) \) will be most similar to \( P(B) \) when the probability of reading plates of different vehicles accurately is similar. This statistical analysis provides insight into the likely magnitude of \( P(B \mid A) \) but does not provide a practical method of estimating \( OD(AB) \).

Although \( M(AB) \) can be calculated and \( P(A) \) and \( P(B) \) estimated, the values of \( \mu \) and \( \sigma^2 \) are unknown.
Appendix D: Fuzzy matching

This report has concentrated on deriving correction factors for OD estimation equations based on all-digit matches between licence plates read by ANPR systems. However, to illustrate the effect on the magnitude of the correction factor $C(AB)$ of using fuzzy matching, the terms in Equation (3) for journeys from site 2 to site 5 in the am peak have been derived using fuzzy matching. Table D.1 shows ‘Fuzzy $M(AB)$’ and ‘Fuzzy $MD(AB)$’, the number of ‘all-digit plus fuzzy’ matches between the ANPR plates at sites A and B and between the manually extracted plates at site A and the ANPR plates at site B respectively alongside the number of all-digit matched plates ($M(AB)$). If all the errors in the ANPR plates were confined to one character, ‘Fuzzy $M(AB)$’ and ‘Fuzzy $MD(AB)$’ would have been expected to be equal and their ratio ‘Fuzzy $X$’ equal to unity, reducing the variation in $P(B|A)$ (which equals $X/P(A)$).

Table D.1 shows that all-digit plus fuzzy matching increases the apparent number of matches by almost 10% on average and that if ‘Fuzzy $MD(AB)$’ is used ‘Fuzzy $F(B|A)$’ and ‘Fuzzy $C(AB)$’ were still very variable as shown in Figure D.1. Table D.2 shows that the estimate of ‘Fuzzy $C(AB)$’ changes by over 6%. As $R(A)$ and $R(B)$ will not have changed for these samples, fuzzy matching would have changed the apparent $OD(AB)$ by almost 3% (since $OD(AB)/M(AB) = R(A)R(B)C(AB)$), as shown by the last row of Table D.2. These results should be treated with caution as individual values of ‘Fuzzy $P(A)$’, ‘Fuzzy $Acc(A)$’ and ‘Fuzzy $Acc(B)$’ have not been used. This limited analysis shows that using fuzzy matching is unlikely to reduce the errors in the estimation process, although further work is necessary to estimate the ‘fuzzy’ values more accurately. However, the need for a correction factor has been confirmed.

Table D.1 Calculation of $C(AB)$ for 5-minute intervals in all lanes using fuzzy matching

| AM peak period | Manual count | All digit matches ($M(AB)$) | All digit + fuzzy matches ($Fuzzy M(AB)$) | All digit manual with ANPR matches ($Fuzzy MD(AB)$) | Fuzzy $X$ | Fuzzy $P(A)$ | Fuzzy $Acc(A)$ | Fuzzy $Acc(B)$ | Fuzzy $F(B|A)$ | Fuzzy $C(AB)$ |
|----------------|--------------|-----------------------------|------------------------------------------|---------------------------------------------|---------|-------------|---------------|---------------|----------------|----------------|
| 2 to 5         | 8:20-8:25    | 302                         | 28                                       | 28                                          | 27      | 27          | 1.04          | 0.88          | 0.96           | 0.95           | 1.17             | 1.07             |
| 2 to 5         | 8:25-8:30    | 381                         | 39                                       | 46                                          | 42      | 48          | 0.96          | 0.88          | 0.96           | 0.95           | 1.09             | 0.99             |
| 2 to 5         | 8:30-8:35    | 281                         | 34                                       | 37                                          | 37      | 40          | 0.93          | 0.88          | 0.96           | 0.95           | 1.05             | 0.96             |
| 2 to 5         | 8:35-8:40    | 272                         | 24                                       | 24                                          | 25      | 26          | 0.92          | 0.88          | 0.96           | 0.95           | 1.05             | 0.95             |
| 5 from 2       | 8:40-8:45    | 259                         | 33                                       | 38                                          | 35      | 43          | 0.88          | 0.88          | 0.96           | 0.95           | 1.14             | 1.04             |
| 5 from 2       | 8:45-8:50    | 263                         | 37                                       | 42                                          | 34      | 42          | 1.00          | 0.88          | 0.96           | 0.95           | 1.35             | 1.23             |
| 5 from 2       | 8:50-8:55    | 232                         | 28                                       | 31                                          | 31      | 31          | 1.00          | 0.88          | 0.96           | 0.95           | 1.51             | 1.38             |
| 5 from 2       | 8:55-9:00    | 265                         | 35                                       | 37                                          | 37      | 42          | 0.88          | 0.88          | 0.96           | 0.95           | 1.10             | 1.01             |
| Numerical average (unweighted) | 282     | 32                         | 35                                       | 34                                          | 37      | 0.95        | 0.88          | 0.95          | 0.95           | 1.18           | 1.08             |
| Standard deviation | |                |                                          |                                             |         |             |               |               |                |               | 0.06            | 0.16            | 0.15            |

Table D.2 Comparison of estimated matches and correction factors

<table>
<thead>
<tr>
<th></th>
<th>All digit</th>
<th>All-digit + Fuzzy</th>
<th>% increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M(AB)$</td>
<td>32.250</td>
<td>35.375</td>
<td>9.69%</td>
</tr>
<tr>
<td>$X$</td>
<td>0.951*</td>
<td>0.954</td>
<td>0.26%</td>
</tr>
<tr>
<td>$C(AB)$</td>
<td>1.013*</td>
<td>1.079</td>
<td>6.54%</td>
</tr>
<tr>
<td>$M(AB)/C(AB)$</td>
<td>31.84</td>
<td>32.78</td>
<td>2.98%</td>
</tr>
</tbody>
</table>

* Estimated from a sample which excludes three datasets
Figure D.1 Variability of ‘Fuzzy F(B|A)’ and ‘Fuzzy C(AB)’ in Table D.1 for intervals 100 (8:20-8:25) to 107 (8:55-9:00)
Abstract

This report describes research into the estimation of the O/D throughput (number of vehicles travelling between two locations) using licence plate data collected using Automatic Number-Plate Recognition (ANPR) systems. The objectives were to recommend a method of estimating O/D throughput and its accuracy (confidence limits) and to recommend good practice for the implementation of the recommended method. The approach used was to collect information from a programme of video filming and measurements at six locations where ANPR systems were installed, and to study the effect of factors such as the traffic conditions and the vehicle characteristics on the accuracy of licence plates read and the number of plates matched. An analytical approach was also used, based on statistical theory, to inform the research by establishing theoretical relationships between the O/D throughput and number of plates matched. The final recommended method related the O/D throughput to the number of matches, the number of plates read at each site and a correction factor, to be derived from a validated reference set of data extracted from video filming. A free-standing document included as an appendix to the report details the recommended method and its implementation.

Related publications

SR757 OECD seminar on road traffic information 23-25 September 1980. 1983 (£35)
SR632 Inexpensive aerial photography for highway engineering and traffic studies by W Heath. 1980 (£20)
LR877 The choice of route, mode, origin and destination by calculation and simulation by D I Robertson and J V Kennedy. 1979 (£20)

Prices current at January 2004

For further details of these and all other TRL publications, telephone Publication Sales on 01344 770783, or visit TRL on the Internet at wwwTRL.co.uk.