Development of network fitting and data handling procedures for a national SCANNER database

N Anderson, C Turnbull, D Wright and A Wright
Development of network fitting and data handling procedures for a national SCANNER database

by N Anderson, C Turnbull, D Wright and A Wright (TRL)

Prepared for: Project Record: Development of network fitting and data handling procedures for a national SCANNER database

Client: Department for Transport
Richard Campbell

Copyright Transport Research Laboratory March 2009

This Published Report has been prepared for Department for Transport. Published Project Reports are written primarily for the Client rather than for a general audience and are published with the Client’s approval.

The views expressed are those of the author(s) and not necessarily those of the Department for Transport.

<table>
<thead>
<tr>
<th>Name</th>
<th>Date Approved</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project Manager</strong></td>
<td>Alex Wright</td>
</tr>
<tr>
<td><strong>Technical Referee</strong></td>
<td>Barry Cleave</td>
</tr>
</tbody>
</table>
When purchased in hard copy, this publication is printed on paper that is FSC (Forest Stewardship Council) registered and TCF ( Totally Chlorine Free) registered.
# Contents

**Executive summary**  
3

**Abstract**  
5

1** Introduction**  
7

2** Establishing the National Framework**  
8

3** Developing Data Fitting Rules**  
9

3.1 Test Networks  
9

3.2 Data  
10

3.3 The fitting process  
11

3.4 Outputs  
12

4** Data processing**  
14

4.1 Determining the environment  
14

4.2 Obtaining RCI  
14

4.3 Obtaining NI (formally BVPI)  
15

4.4 Reporting in relation to section and chainage  
17

5** Assessment of the fitting process**  
18

5.1 Manually checking for errors  
18

5.2 Checking coverage of ITN sections  
18

5.3 Counting the unmatchable points, and the effect of adjustable parameters  
19

5.4 Measure of certainty for matched points  
20

5.5 Dual carriageways  
21

6** Impact of fitting upon reporting RCI and NI**  
23

6.1 Different types of error  
23

6.2 Comparison of RCI values from UKPMS and RCI fitted SCANNER data  
25

6.3 Comparison of NI results from UKPMS and ITN fitted SCANNER data.  
27

7** Implementation**  
32

7.1 Overall process for delivering a national SCANNER database  
32

7.2 IT requirements  
32

8** Conclusions and recommendations**  
35

**Acknowledgements**  
36
Appendix A  Establishing the national framework  37
Appendix B  The data fitting process  41
Appendix C  Dual carriageways  48
Appendix D  The scoring process to determine the likelihood of matching  52
Executive summary

The SCANNER survey provides local authorities with an objective assessment of the condition of the (surveyed) local road network. In England SCANNER is the only method used for the calculation of performance indicators for the reporting of condition of the classified road network. The survey therefore offers the clear benefit of objective, accurate and consistent measurement of condition, carried out over the whole classified road network.

Although there is considerable potential for the use of the SCANNER data both locally and nationally, centrally the Department for Transport (DfT) sees only the summary data reported as a National Indicator (NI) for each Local Authority, which is a single number indicating the length of the Authority’s highway network that is classified as being in poor condition.

By storing the SCANNER data centrally the DfT would have access to a powerful database of network condition. Such a database would provide the potential for the Department to undertake a wide range of analyses. These could include assessing the extent of maintenance and improvement being undertaken on the network, classifying the extent to which specific defects exist on the network, or investigating the effects of changing the calculation of the National Indicator (e.g. to improve its ability to correctly identify poor lengths).

However, to provide such a central database there is a requirement to develop methods to take the data from the different local authority networks, surveys and survey contractors and fit this data to a single defined reference framework, against which each data set can be aligned, displayed and reported. There is then a need to develop procedures for the processing of the data so that additional information can be obtained (for example Road Condition Index (RCI) or NI values).

TRL was therefore commissioned by DfT to produce a methodology for the fitting of SCANNER data to a national reference framework (road network) to allow data to be viewed, processed and analysed on a national (rather than local) level. The results of this work are presented in this report.

Only a few networks exist that could be used as a national framework for referencing the SCANNER data. Of these, the Ordnance Survey MasterMap ITN is recommended as the most appropriate network for use in a national SCANNER database.

Processes have been developed for fitting the SCANNER data to the national framework. The fitting is carried out in three stages. Initial pre-processing of the delivered SCANNER data converts the data into a format suitable for fitting. This pre-processing extracts the location data from the SCANNER survey data for subsequent fitting to the network. Fitting algorithms are then applied to the location data to identify the appropriate Ordnance Survey ITN (ITN) section to which the SCANNER data should be fitted. This process delivers the SCANNER data, but now referenced to the ITN. In this project the fitting process has been implemented in a standalone system as an “executable product”. It has been found that this delivers a high level of efficiency for the fitting process. However, the algorithms could be implemented in a conventional database if required. This could result in a significant increase in the processing time required for the fitting process.

Tests of the fitting have shown that the level of performance is high. Manual assessment of a random sample of data from a test network identified less than 0.5% of mismatched points, with less than 0.05% matched to roads of the wrong classification. Less than (approximately) 1% of SCANNER points were identified as “unmatchable”.

One area where the performance of the algorithms could be improved is in the fitting of SCANNER data collected on dual carriageway roads, where the data may be fitted to the wrong side of the dual carriageway. Algorithms have been suggested that may
assist in improving the fitting of SCANNER data collected on dual carriageways. However, it is not recommended that these be implemented without further development. Furthermore, assessments using an urban test dataset, which was considered likely to give the worst case, showed that less that 0.1% of the data had been incorrectly fitted as a result of fitting to the wrong side of a dual carriageway.

Processing of the fitted SCANNER data can be carried out using the functionality provided by standard GIS and database systems. However, a number of external data sources are required in order to obtain measures such as the SCANNER RCI and National Indicators (NI). Data on the environment (urban/rural) can be provided from the national definition without significant effect on the results of the processing. Information on the classification can be obtained from the national framework. However, the recommended national framework (ITN) does not provide the facility to distinguish between C and U roads, which could lead to errors where local authorities have commissioned SCANNER surveys on these roads. It is suggested that SCANNER contractors be asked to provide the classification of the roads surveyed when providing the SCANNER data (for example by providing data from A, B, C and U roads in separate data files).

Following the completion of processing it is straightforward to apply standard rules for the calculation of National Indicators. The project has undertaken this calculation for two test networks, one primarily urban, the other primarily rural. The values of National Indicators NI168 (A roads) and NI169 (non-principal classified roads) were calculated for these test networks using both the standard UKPMS approach and using the fitted SCANNER database. Similar values were obtained for the indicators for both methods.

An implementation route for the database has been proposed, which recommends the use of dedicated software tools for the pre-processing and fitting of the SCANNER data, followed by the use of database and GIS facilities for the processing, display and reporting of the data. It is felt that the system could easily be implemented using typical office IT equipment. At the commencement of this project it had been suggested that there would be a need to process the SCANNER data from the delivered 10m lengths into 100m lengths in order to reduce storage space and optimise the processing time. However, the project has found that this should not be necessary.
Abstract

The SCANNER survey provides local authorities with an objective assessment of the condition of the (surveyed) local road network. Although there is considerable potential for the use of the SCANNER data both locally and nationally, centrally the DfT sees only the summary data reported as a National Indicator for each Local Authority. By storing the SCANNER data centrally the DfT would have access to a powerful database of network condition. Such a database would provide the potential to undertake a wide range of analyses. However, to provide such a central database there is a requirement to develop methods to take the data from the different local authority networks, surveys and survey contractors and fit this data to a single defined reference framework, against which each data set can be aligned, displayed and reported. There is then a need to develop procedures for the processing of the data so that additional information can be obtained (for example Road Condition Index (RCI) or NI values).

TRL was therefore commissioned by DfT to produce a methodology for the fitting of SCANNER data to a national reference framework (road network) to allow data to be viewed, processed and analysed on a national (rather than local) level. The results of this work are presented in this report.
1 Introduction

The SCANNER survey provides local authorities with an objective assessment of the condition of the (surveyed) local road network and in England is the only method used for the calculation of performance indicators for the reporting of condition of the classified road network. The survey therefore offers the clear benefit of objective, accurate and consistent measurement of condition, carried out over the whole classified road network.

Although there is considerable potential for the use of the SCANNER data both locally and nationally, centrally the DfT sees only the summary data reported as an indicator (formally the Best Value Performance Indicator – BVPI, now National Indicator - NI) for each Local Authority, which is a single number indicating the length of the Authority’s highway network that is classified as being in poor condition.

By storing the SCANNER data centrally the DfT would have access to a powerful database of network condition. Such a database would provide the potential for the Department to undertake a wide range of analyses. These could include assessing the extent of maintenance and improvement being undertaken on the network, classifying the extent to which specific defects exist on the network, or investigating the effects of changing the calculation of a National Indicator (e.g. to improve its ability to correctly identify poor lengths). However, to provide such a central database there is a requirement to develop methods to take the data from the different local authority networks, surveys and survey contractors and fit this data to a single defined reference framework, against which each data set can be aligned, displayed and reported. There is then a need to develop procedures for the processing of the data so that additional information can be obtained (for example Road Condition Index (RCI) or NI values).

TRL was therefore commissioned by DfT to produce a methodology for the fitting of SCANNER data to a national reference framework (road network) to allow data to be viewed, processed and analysed on a national (rather than local) level.

The report presents the work carried out in development of the development of network fitting and data handling procedures for a national SCANNER database. The key technical stages of this task included:

- Establishing the national framework – discussed in section 2
- Developing and demonstrating data fitting rules – discussed in section 3
- Establishing the data processing methods – discussed in section 4
- Assessment of the performance of the fitting process – discussed in section 5
- Assessment of the effect of fitting on data reporting – discussed in section 6

Finally, in Section 7 the steps that are required for larger scale implementation of the database are discussed.


2 Establishing the National Framework

The vision for a national SCANNER database is to have the SCANNER data stored centrally, with the ability to present the data to central users for analysis of network condition. To achieve this, the user requires a facility to select the parts of the data that require display and analysis. There is therefore need for a system of referencing to be adopted for all of the data. SCANNER data is collected, processed and delivered with reference to Local Authority defined networks. These networks, which are defined locally by each local authority, break the local network into sections. The network is “live” in that it may be changed by the authority to accommodate local needs (e.g. changes to priorities, or to simplify the layout). At the commencement of this project it was accepted that it would be impractical to obtain and maintain copies of each local authority network definition within the central database. Instead, it was decided that the SCANNER data would be “fitted” to a national network definition. The first phase of this project, establishing the national framework, therefore determined a suitable network to which to fit all of the SCANNER data.

A review was undertaken of the appropriate available networks to determine which would be the most suitable for use as the basis of a national SCANNER database. The results of this review were delivered in an interim report in March 2008. This report is reproduced in Appendix A, and a summary of the conclusions are presented within this section.

The review identified only a few potential national networks. In fact the majority of potential network are based either partly or wholly on the Ordnance Survey MasterMap ITN (Integrated Transport Network). However, in addition to Ordnance Survey, it was found that TeleAtlas and Navteq also hold similar networks. These two companies provide the data used by almost all of the satellite navigation systems currently available.

A comparison of the three key networks did not show any particular network to have remarkable strengths over the others (as discussed in Appendix A). Additional factors were therefore considered, and the Ordnance Survey MasterMap ITN was recommended for the following reasons:

- The OS ITN contains as much data as the other two datasets, but where there are differences these favour the use of the OS ITN. In particular the MasterMap Topology layer – which gives an accurate polygon representation of the roads – may assist in fitting where there is a high concentration of roads, and could improve the accuracy of the fitting process.
- The OS ITN is used throughout Government for mapping and displaying data. It would make sense to use the same referencing system for the National SCANNER Database to allow for an easy exchange of information.
- DfT already have a license for the OS data. Whilst this may not be of relevance if there were a clear advantage to using one of the other datasets, given that they all are of similar suitability it would not be sensible to spend money to procure another dataset.
- The OS ITN is updated more frequently than the other two networks.
- None of the networks include C road designations.
3 Developing Data Fitting Rules

SCANNER data consists of pavement condition measurements, such as rutting and cracking, recorded during the survey and processed to provide summary data reported over 10m longitudinal lengths. As noted in Section 2, SCANNER network referencing is based on the traditional approach of road sections, which each describe particular lengths such as “the A123 from the junction with the B123 to the junction with B234”. Hence the SCANNER data is delivered as the condition at each position along each section. However, the SCANNER survey equipment also records the geographical location of the survey vehicle (using GPS), which is processed to provide OSGR\(^1\) co-ordinates reported at 10m intervals, these measurements being well aligned with the condition data reported at the same intervals. The SCANNER condition data can therefore be fitted to an alternative network, using the OSGR co-ordinates provided within the SCANNER dataset, providing that the external network is also referenced to these co-ordinates.

The selected national network (Ordnance Survey MasterMap ITN) is also described in terms of “sections” (which are not related in any way to the Local Authority’s definition), and also in terms of OSGR co-ordinates. The second phase of this project therefore concentrated on establishing procedures (or rules) to translate (or “fit”) the SCANNER data to the Ordnance Survey ITN, using the OSGR co-ordinates. The aim of this process was to deliver SCANNER data that was no longer reported in relation to the original local authority sections, but to the ITN sections, using the OSGRs and the cross-referencing method.

3.1 Test Networks

Two test networks were selected for the development of the fitting process. These encompassed the Birmingham and West Sussex local authority road networks. These were selected because they contained areas representative of urban, rural and mixed networks. The Birmingham dataset was approximately 1400km long, the West Sussex data set was approximately 2150km long. An ITN snapshot covering the regions of the survey data was made available – see Figure 1.

![Figure 1: GIS representations of the ITN road networks of Birmingham (left) and West Sussex (Right)](image)

\(^1\) Ordnance Survey Grid Reference
3.2 Data

SCANNER data is provided by SCANNER survey contractors as BCD (Base Condition Data) files. This is a pre-defined file format (known as HMDIF) that breaks the survey data into a series of sections, within which the SCANNER data is reported at 10m intervals. BCD files may contain one or more sections. Some SCANNER survey contractors provide the data from each local authority in which they have undertaken surveys as a single BCD file. Other contractors provide multiple BCD files, each containing the data from a single survey run of the survey vehicle. Regardless of the method of delivery (single or multiple files), SCANNER data consists of a series of points within each section, which are typically (but not necessarily) provided in the order in which they were surveyed. The sections are often reported within individual data files in the order over which they were surveyed. However, it is not a requirement of SCANNER that the data be delivered in any particular order. Therefore the fitting process should not make any assumptions about the order in which the SCANNER data is provided, as there are no guarantees regarding this.

For simplicity the SCANNER data used in developing the fitting rules, which was provided in multiple BCD files for the test networks (section 3.1), was pre-processed to deliver a single data file containing all of the SCANNER data. Before the commencement of this project it had been suggested that there would be a need to process the SCANNER data from the delivered 10m lengths into 100m lengths. It was felt that this would be necessary in order to reduce storage space and optimise the processing time. However, it was noted that this would bring the disadvantages of reduction in detail and possible disagreements between the reports (such as NIs) obtained from the national database and those obtained locally by Local Authorities. However, initial stages of the project used the 10m data for the development of the fitting algorithms and it soon became apparent that efficient fitting processes could be developed with this data. Hence it would not be necessary to reduce the data to 100m lengths in order to fit the data to the national network in an acceptable length of time. An assessment of the space required to store the data also showed that a database of 10m values could easily be accommodated by standard IT systems. Therefore, there has been no attempt made in this project to develop a process that uses a reduced (100m length) dataset. However, initial stages of the project used the 10m data for the development of the fitting algorithms and it soon became apparent that efficient fitting processes could be developed with this data. Hence it would not be necessary to reduce the data to 100m lengths in order to fit the data to the national network in an acceptable length of time. An assessment of the space required to store the data also showed that a database of 10m values could easily be accommodated by standard IT systems. Therefore, there has been no attempt made in this project to develop a process that uses a reduced (100m length) dataset. However, it should be noted that the pre-processing stage (to generate single SCANNER data files for each authority) does assist in reducing the quantity of file handling required by the fitting process. Hence it is proposed that this method be used to simplify data handling in the final implementation of the database – see section 7.

For the purpose of developing the fitting rules the location information in the SCANNER data was processed into a Shapefile\(^2\) containing the original local authority section information and the geographical data.

ITN consists of a shape file containing a series of roads. Each length of road is described by a TOID\(^3\). Roads bend and curve, hence each TOID is made up of many road sections. Each road section is typically of length 5 – 10m. Geographical (OSGR) location information is also included in the ITN shape file.

\(^2\) A shapefile is a geospatial vector data format commonly used by GIS software. Shapefiles can be used to spatially describe the geometry of features including points, lines, and polygons. The format provides the facility for features to have attributes, such as the name.

\(^3\) Topographic IDentifier. This is a unique reference identifier used to identify features. Every object in the Ordnance Survey MasterMap ITN has its own TOID, which could describe a polygon, line, point etc.
3.3 The fitting process

The fitting process is required to extract each point from the SCANNER data and – using the OSGR co-ordinate of each point – search the ITN and identify the correct ITN section in which each point lies. After fitting a point to an ITN section, the precise location of the point on the ITN section can be estimated by a method of projection. Once the fitting is complete it will be possible to assess road condition in relation to the ITN sections defined within the national network.

A number of challenges became immediately apparent for successful development of the fitting process:

- The ITN network is very large. A suitable approach needed to be developed so that the process of identifying the appropriate ITN section was efficient.
- The ITN network is complex. SCANNER OSGR data points could lie close to many potential ITN sections, complicating the decision on the correct ITN section in which the point should lie. The process needed to be able to determine which of these sections would be correct. It was therefore decided that a process would be required to indicate the approximate level of confidence achieved in fitting each SCANNER data point, therefore allowing for points to be left not fitted where the confidence was low.
- The information contained with the ITN is not complete or exhaustive, and not necessarily well matched by the Local Authority definition. For example the direction of traffic flow on a road is not included in the ITN, but would probably be required in order to make good use of the SCANNER data (e.g. in calculating RCIs from multiple survey years). Therefore there was a need to develop processes that could accommodate these requirements.

The following sections briefly describe the fitting process that has been developed to address the above challenges. The process consists of four steps:

1. Inspection of the ITN and indexing of the roads therein. This index is used to optimise (improve the efficiency/speed of) the fitting process – section 3.3.1.
2. Initial analysis of each point within the SCANNER data to identify any potential ITN road sections to which each SCANNER could match – section 3.3.2.
3. Projection of the SCANNER points onto the potential ITN sections. Assessing the likelihood of each potential match (as a score), rejecting matches that lie below a certain threshold – section 3.3.3.
4. Conducting additional processing to compensate for known and avoidable errors – section 3.3.4.

These steps are described briefly in the following sections. A more detailed description of the process is presented in Appendix B.

3.3.1 Indexing

In order to match SCANNER data points to roads within the ITN, each point must be tested for compatibility against a number of ITN road sections. With an un-optimised system there is the potential that each SCANNER point would have to be compared with a large proportion of the ITN road sections in order to determine which ITN road section the SCANNER points sit within. Ideally, the number of ITN road sections tested should be minimised to increase the efficiency of the algorithm. In the process developed in this project, optimisation is achieved by indexing all the roads in the ITN into a three-dimensional array, where OSGR Eastings and Northings make up the first two dimensions and the direction of travel (bearing) is the third. Each element in the array is referred to as a segment and each segment may contain a number of ITN road sections.
The concept behind the indexing is that the number of comparisons with the ITN network can be reduced by efficiently isolating the part of the ITN (the road sections) that fall close to the SCANNER point under consideration. This is done by indexing on Easting and Northing. A further restriction is then applied that describes the direction of travel of the ITN road section (the third dimension in the index) for comparison with the direction of travel of the SCANNER data. This further reduces the number of ITN road sections that have to be compared with each SCANNER data point.

### 3.3.2 Initial matching of SCANNER points to the ITN

Following the indexing of the ITN data the SCANNER data points can be compared with the indexed ITN data to identify the potential ITN road sections within which the SCANNER data point may fall. In the process proposed in this work the algorithm identifies the array segments in which the SCANNER data point sits and also the adjacent array segments. All ITN road sections within these array segments are potential matching ITN road sections.

### 3.3.3 Projection and determining likelihood of match

For each potential road section in the array segment to which the SCANNER data point has been matched, the closest point on that road section to the SCANNER data point is calculated – this point on the road section is referred to as the projected point. The distance between the projected point and the original SCANNER data point, as well as the difference between the direction of the road section and direction of travel of the SCANNER vehicle at the survey point, is used to calculate the likelihood that the projected point is a good match. See Appendix B and Appendix D for more details.

After all road sections in the array segment have been analysed, the road section containing the most likely projected point (i.e. having the lowest score) is selected as the final matching ITN road section.

### 3.3.4 Additional processing

The above process was found to successfully match the “easier” of the SCANNER data points to the correct ITN road section (i.e. points where there were a limited number of possible matches, no parallel roads, etc). However, visual assessment of the fitting process identified some issues with the matching in more complicated locations. Hence additional processing steps were required to improve the process. In particular the presence of slip roads, which occupy almost the same space and have similar directionality to main carriageway roads, can cause SCANNER data points obtained on a main road to be erroneously matched. It suggested that such errors could be reduced by looking for discontinuities in the fitted points. Where discontinuities are present the process checks the surrounding points and assesses the probability that an error has occurred, action can then be taken to remedy the error and match points to a road where no discontinuities happen. For more details of this process, see Appendix B, Section B.5.

### 3.4 Outputs

After the fitting process is complete, the resulting dataset consists of the original SCANNER data points, translated to appear on the ITN road section to which they have been fitted. The impact of the fitting algorithm upon the data set can be visualised using GIS mapping software. Figure 2 shows an example set of SCANNER data (green) shifted onto to the ITN network (red) after application of the fitting rules.
For network level processing it will only be necessary to determine which ITN road section to which each SCANNER data point has been fitted. However, because of the projection process, it is also possible to estimate the position along the ITN road section to which the SCANNER data has been fitted. This additional level of detail enables good visualisation of the data, for example when displayed in a GIS, because it enables the SCANNER data to be visualised in a way that would be expected by the user – i.e. in a manner that appears to “travel” along the road. However, this representation should not be confused with the capability to process SCANNER data fitted to the ITN as if it was still related to section and chainage along the length of a road, this is discussed in more detail in section 4.3.

Figure 2: SCANNER data in its original and ITN matched locations at a large dual carriageway junction in Birmingham.
4 Data processing

4.1 Determining the environment

A key requirement of the central analysis of the SCANNER database will be the calculation of RCI and NI values. To obtain these data it is required that:

- The SCANNER data has been fitted to the ITN
- An urban/rural attribute has been set for each fitted SCANNER data point
- The survey year is known for each fitted SCANNER data point
- The road classification is known for each fitted SCANNER data point
- The road type (single / dual carriageway) is known for each fitted SCANNER data point

The calculation therefore requires some work subsequent to the completion of the fitting process. In particular the allocation of urban/rural and classification attributes to the data is required because the rules and parameters of the SCANNER RCI are based on environment (urban/rural) and classification.

The determination of the urban/rural environment is easily achieved by comparing the fitted SCANNER data to known urban/rural definitions provided geographically, for example within a centrally defined shape file. The shape file may be compared with the fitted data within a GIS to apply the urban/rural attributes to the SCANNER data points.

Clearly, knowledge of survey year is simply accommodated within the database by setting the appropriate attribute when the SCANNER data is initially fitted and loaded.

The classification may be taken from the ITN data for A and B class roads. However there are no means within the ITN to identify C class roads. Two possible methods are proposed to identify C class roads:

- If SCANNER data was initially supplied with each road classification in separate files, then the road class could be recorded at that time and passed through the pre-processing to allow selection of data.
- Alternatively, classification can be based on the SCANNER data itself. Assuming that SCANNER surveys are only carried out on A, B and C roads (the current compulsory requirement), any lengths for which SCANNER data was provided that were not recorded in the ITN as A or B roads must be C roads.

There are obviously limitations to both methods. In either case, C road sections will only be identified once SCANNER data had been provided for those sections. It will be difficult to determine the percentage coverage achieved on the C road network. For the second method there is a risk that data collected by SCANNER on unclassified roads (which are carried out in some Local Authorities) are wrongly labelled as C road data and included in RCI/BVPI calculations.

The type of road is available from the ITN, in which the road can be defined as, amongst others, single or dual carriageway.

4.2 Obtaining RCI

IN UKPMS the RCI is calculated by applying the appropriate UKPMS weighting set to the observations recorded. Within the weighting set, a number of curves are defined. These are used to rate each observation. There are also confidence factors which further scale the results. Both curves and confidence factors vary depending on road class and urban/rural classification. Observations are also grouped into families and
when producing the RCI, the highest rated value from each family is carried forward and contributes to the final RCI.

The process of obtaining the RCI from the fitted data would be as follows:

- From the fitted points data, select the first point.
- Determine the road classification and environment (see section 4.1).
- From the attributes data corresponding to the fitted point, select the appropriate parameters for use in the calculation (rutting, profile, texture, cracking). Apply the weightings given in the current version of the Rules and Parameters to produce a weighted score for each parameter.
- Combined the weighted scores to calculate the RCI.
- Repeat for all points.

Following the calculation of RCI values, it is envisaged that the data would be used to produce a colour coded plot of the points, such as the Red Amber Green (RAG) map of RCI values commonly produced from UKPMS systems. With a suitable choice of size for the points represented in the GIS, the appearance of a continuous colour banding along the roads can be achieved – see Figure 3.

![Figure 3: An example of an RCI map showing road condition. The map is visualised in a GIS system.](image.png)

### 4.3 Obtaining NI (formally BVPI)

Indicators will form a key output of the national database. Until 2008, the key SCANNER indicators were the Best Value Performance Indicators for principal roads (BV223) and Non-Principal Classified roads (BV224a). Whilst these BVPIs have now been withdrawn, they have been replaced by National Indicators NI168 and NI169...
respectively. These Indicators are effectively a percentage of the road that has an RCI of over 100 (the band defined as ‘Plan maintenance soon’). NI168 is produced for A roads and NI169 for B and C roads (the results for the B and C roads are combined based on the length of each road type in the network). Technical notes on the production of BVPIs were produced annually until 2007/08 (TN36 and 37 respectively). At the time of carrying out the work technical notes had not been published for the NIs. This work has assumed that the calculation of the new NIs will not differ from that used for the calculation of the BVPIs, and hence the term “NI” used in this report is interchangeable with the term “BVPI”. It should not be assumed that values of NI/BVPI presented herein are representative of the values that would be obtained by the named local authorities when reporting their performance indicators.

NI values can be used to demonstrate the relative performance of roads both within and across local authorities. NIs can be calculated on selected ITN sections using a GIS. However, the delivery of NI data from a GIS requires the implementation of the appropriate calculation/reporting rules within the GIS.

It is envisaged that a user would select the sections to be used in calculating the NI by area, road or other suitable means. The set of TOIDs would then be passed to the NI reporting process. At this stage the user would also select the year(s) over which SCANNER data is to be selected for use in the calculation of the NI.

The following steps would then be carried out:

- The reporting process will identify all points from the relevant years falling within the selected sections, including the RCI values for the selected points (see 4.3.1).
- The coverage for each TOID will be calculated. This involves counting the number of points that are fitted to each TOID. If the number of points equate to 50% or more of the TOID length, it is considered to be covered. This step ensures that odd data points that may incorrectly have been fitted to a TOID are not taken to represent a complete survey of that TOID.
- Taking account of whether a section is dual or single carriageway, the most recent sets of data from the previous step would be selected, either one set for dual carriageways or one set for each direction on single carriageways.
- The process will calculate the number of points for which the RCI value exceeds the NI thresholds, and convert this to length, based on the assumption that each point represents 10m on the network.
- The total length over which the NI is being obtained will be calculated, based on the sum of the lengths of the selected ITN sections.
- The process will report the NI values, by road class, as appropriate. It is suggested that NI data be produced in a report format similar to that delivered in the UKPMS NI (formally BVPI) report. The reporting would either contain results for A, B and C roads in one report, or split between A roads and B & C roads.

4.3.1 Dealing with year on year data

By use of the above method, it should be possible to use data from as many years as required. The filtering will select the most recent set of data (within the date range provided by the user) to be used for each TOID. It will also include data for each direction for single carriageway TOIDS (it is not unusual for one direction to be surveyed one year and the other in a different year).
4.4 Reporting in relation to section and chainage

The above sections discuss the calculation of RCI values for selected lengths of the network and display on a GIS, and the subsequent calculation of NIs. For more detailed analysis of the data it may be desirable to provide other viewing and reporting facilities. An obvious technique is display on a graph (for example RCI vs. chainage). This would reflect the traditional reporting of road condition data on a distance basis, i.e. using road sections and chainage to locate the position of a data point on a road. Presently such an approach is not possible following fitting of the SCANNER data to the ITN network, because of the limitations of section referencing within the ITN.

After being matched to the ITN, the SCANNER condition data is stored in a geographical format rather than being referenced to roads via a road network labelling system. Each point is located via an OSGR co-ordinate pair and the road section to which it was matched. Each road section is part of a larger grouping of sections known as a TOID (see section 3). The position of road sections within a TOID is not recorded in the ITN, and as such the chainage of a SCANNER data point within a TOID may be complex to compute. In addition, the SCANNER data points are recorded in a database rather than a sequential file format. As such there is no guarantee that the points are in order of collection, a fact which would make the generation of a distance-based condition data report more complicated still.

A distance-based reporting system from the ITN-fitted SCANNER data would require the development of an automated methodology which sought to locate data points within individual TOIDs. This methodology would need to determine the locations of road sections within each TOID and sum their lengths until the required data point was reached. Vehicle direction of travel would allow such a system to determine the side of the road on which a point belongs, although lane designations would not be possible.

The development of such a methodology is outside the scope of the fitting algorithm, but may be a desirable aspect of the final implementation of the system.
Assessment of the fitting process

Assessment of the success of the fitting process was carried out using the test networks. A number of techniques were applied:

- Manually checking for errors
- Checking the coverage of ITN sections
- Counting the unmatchable points, and assessing the effect of adjustable parameters
- Measure of certainty for matched points

5.1 Manually checking for errors

Manual checking was carried out by exporting a randomly selected 5% of the output data to a test file and visually inspecting the successful matching to the ITN on a GIS.

In the sample selected from West Sussex, the data from 98 SCANNER sections was assessed, containing 10482 individual points (around 100km of data). The checking process consisted of displaying the points in the GIS system and checking that the points had been matched to the correct TOID in the ITN. A count was kept of the number of points which had been matched to the wrong TOID, and also those where the mismatch was to a TOID that was of a different road class (i.e. A, B, C, Unclassified). This latter test was included because matching to an incorrect class of road could have a significant effect on the calculation of RCI and NI.

In the data set checked, there were a total of 33 (0.31%) mismatched points. These generally occurred in locations near the start or end of a SCANNER section, particularly on the entry/exits of roundabouts or where the survey vehicle was turning into/out of the road. Of the mismatched points, only 2 (0.02%) were matched to roads of a different class.

This error rate is unlikely to cause any noticeable change in RCI and/or NI results at either a local or national level. Occasionally these fitting ‘errors’ may actually have arisen from the SCANNER data being more representative of the vehicle’s location than the ITN – i.e. in cases where the survey vehicle is turning from one road to another. These can hardly be categorised as errors arising from the fitting process.

5.2 Checking coverage of ITN sections

The coverage check is based on the assumption that ITN TOIDs would typically be fully “covered” by the SCANNER survey. This means that we would expect SCANNER data to be provided for the length of most ITN TOIDs, and not to terminate part way through. Hence a TOID with only a few SCANNER points allocated to it may indicate that those points have been allocated to the wrong TOID. A method was applied that classified an ITN TOID as “covered” if at least 50% of the TOID was covered by SCANNER observations. This then enabled us to identify the number of SCANNER points that had been allocated to TOIDs not classed as covered. It was felt that this would give some quantitative measure of the number of points incorrectly matched.

Within the West Sussex dataset the number of points allocated to TOIDs that had only partial coverage was 400, out of a total of more than 200,000 points. Hence a little under 0.2% of SCANNER points were allocated to partially covered TOIDs.
5.3 Counting the unmatchable points, and the effect of adjustable parameters

An unmatchable SCANNER point is defined as one for which no ITN road section can be found onto which the point can be projected. Actually we define the inability to match to be an inability to obtain a likelihood score (see section 3.3.3) of lower than a predefined threshold. The number of unmatchable points can therefore be reduced by modifying the parameters used in the matching algorithm - increasing the size of the ITN index angle or raising the maximum permitted score. Work was therefore undertaken to assess the number of unmatchable points, and the effect of the parameters on this number.

The investigation found that raising the maximum permitted score would be unwise, as it can result in SCANNER points being matched a long way from the initial position. Therefore, although the number of matching points increased, manual assessment showed that there was an increase in the number SCANNER points fitted to incorrect ITN road sections.

Increasing the size of the ITN index angle reduces the effect of small deviations in the position data on the matching process. In a practical sense this means increasing the number of ITN road sections that each SCANNER data point is checked against for a potential match (see section 3.3.1).

Table 1 demonstrates the effect of varying the ITN index angle. The number of unmatchable points decreases. It appears that the number of errors (quantified by manually checked a random output) cannot be directly related to the change in index angle, with the total number of errors in the dataset being generally consistent. This is because the initial scoring and subsequent post-processing steps (moving averaging – see Appendix B) successfully remove many potential errors.

<table>
<thead>
<tr>
<th>ITN index angle (degrees)</th>
<th>Unmatchable Points (%)</th>
<th>Total Errors in sample data set (U-road matches) – (%)</th>
<th>Processing Speed (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2.58</td>
<td>0.34 (0.03)</td>
<td>26</td>
</tr>
<tr>
<td>10</td>
<td>1.71</td>
<td>0.44 (0.05)</td>
<td>27</td>
</tr>
<tr>
<td>15</td>
<td>1.31</td>
<td>0.25 (0.08)</td>
<td>27</td>
</tr>
<tr>
<td>20</td>
<td>1.09</td>
<td>0.22 (0.05)</td>
<td>28</td>
</tr>
<tr>
<td>25</td>
<td>0.84</td>
<td>0.31 (0.04)</td>
<td>28</td>
</tr>
<tr>
<td>30</td>
<td>0.73</td>
<td>0.05 (0.00)</td>
<td>29</td>
</tr>
<tr>
<td>36</td>
<td>0.57</td>
<td>0.16 (0.00)</td>
<td>29</td>
</tr>
<tr>
<td>45</td>
<td>0.37</td>
<td>0.51 (0.00)</td>
<td>29</td>
</tr>
<tr>
<td>60</td>
<td>0.07</td>
<td>0.26 (0.00)</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 1: Varying the ITN index angle decreases the number of unmatchable points and has little impact on the number of manually-observed errors.

Figure 4 demonstrates how a possible improvement in matching can be obtained when increasing the size of the ITN index angle. Points A, B and C are un-matched SCANNER data points plotted according to their grid position. The line between point B and C defines the “direction of travel” when the survey vehicle was at point A. We find that point A is matched to A’ when the ITN angle of indexing limited to 30°. This is because the angle of the road section on which A’ lies is within 30° of the direction of

4 the maximum permitted score during testing was 50
travel of the vehicle at point A. The angle of the road on which A’’ lies (A’’ being closer
to A and therefore a better match) does not lie within this indexing limit. However, in
this example point C is actually a poor representation of the vehicle position, arising
from GPS error. This results in a bad estimation of the direction of travel of the vehicle
at point A. A better match for point A is found when increasing the size of the ITN
angle indexing to 60, then point A can only then be matched against the road section
on which A’’ lies.

Figure 4: The improvement in SCANNER point matching when increasing the
ITN angle segment size is demonstrated by the projection of point A to points
A’ and A’’

5.4 Measure of certainty for matched points
As described above, the fitting process identifies potential ITN road sections and
calculates a possible score for fitting to each road section. The lowest scoring section is
selected as the matched section. Because there is potential that many potential ITN
road sections could be identified, it was felt that an assessment of the range of scores
arising for each match could provide an indication of the certainty of each match.
Hence a simple measure of certainty was developed to provide an insight into how
likely a match is to be correct, in comparison to the second best match. This is a
percentage measure, which is given by equation 1.

\[
\frac{(\text{Max possible score} - \text{Chosen match score} + \text{Second best score}) \times 100}{2 \times \text{Max possible score}}
\]

For the measure of certainty a value of 100% represents an almost certain match and
0% represents a low level of certainty of the selected match in comparison with the
alternative match.

The distribution of certainties for all matches in the test data sets is shown in Figure 5.
The majority of matches are highly certain – i.e. greater than 80%. There is a small
peak at approximately 60%, this is due to the fact that a reasonable number of
matches were in fact borderline, i.e. where the selected match had only a slightly lower
(better) score than the second best match. Matches of this nature result in a certainty score of just higher than 50%. There are a number of ‘matches’ for which zero was scored. These are disregarded and not considered successful matches between the SCANNER data and ITN.

![Distribution of Certainties](image)

**Figure 5: Distribution of certainties of matched data points**

### 5.5 Dual carriageways

The ITN contains limited information on dual carriageway roads. Each side of a dual carriageway is identified as an “independent” road. However, no information about the direction of flow on the carriageway is provided. The effect of this on fitting is that we are unable to isolate the carriageway using the direction of travel. Because dual carriageway road sections are usually close together and have the same bearing, the fitting process will probably consider both carriageways when attempting to fit SCANNER data points. The difficulties this presents may be increased by errors in the SCANNER position measurement (GPS).

Consider Figure 6, which shows an example of SCANNER and ITN data from central Birmingham. The SCANNER position data (green points) can be seen to drift away from the road in the centre of the figure. This is because there is a short tunnel at this location in Birmingham which has caused an error in the reporting of location. In the middle of the tunnel the SCANNER points obtained on the northbound carriageway coincide with the ITN data for the southbound carriageway. The fitting algorithm matches these points to the southbound carriageway, as the ITN contains no information about the direction of flow on roads.

A method for characterising the flow on the ITN network was developed and tested on the available test networks. The methodology employs a system for finding ‘partner’ dual carriageway roads and using this to infer the direction of flow. The method is discussed in more detail in Appendix C. The result of applying this methodology to the drifted points in Figure 6 can be seen in Figure 7, where it is apparent that many of the points have been corrected.

Unfortunately we have found that the algorithm fails to correctly categorise all dual carriageway roads. This is because of an inherent weakness in the algorithm in the identification of dual carriageways lying in a particular direction (east to west in the current method). This weakness can result in correctly-fitted SCANNER data points being moved to the wrong side of a dual carriageway. The method therefore requires further development before being applied to the national SCANNER database.

Although there may be some errors arising in the fitting of dual carriageways, the following observations were also made when assessing the success of the fitting on the test networks:
• A review of the fitted test data in Birmingham revealed that less than 1km of data points had been fitted to the wrong side of a dual carriageway (less than 0.1% of the total).

• It was felt that the Birmingham network was representative of a very challenging dataset, suggesting that the overall error in fitting to dual carriageways on the whole network would probably be less than this figure. Also, the proportion of dual carriageways on the national network is less than in a large city such as Birmingham.

• Improvements have been made to the SCANNER position measurement systems since the date of the SCANNER survey data used to assess the fitting on the Birmingham network. This would assist in improving the fitting to dual carriageways.

Figure 6: The drift of the GPS on the SCANNER survey vehicle is visible on this pass through the Queensway tunnel in Birmingham.

Figure 7: As Figure 6, but with correction for the points which have drifted on to the wrong carriageway.
6 Impact of fitting upon reporting RCI and NI

The following sections assess the impact of fitting errors upon the reporting of local and national road condition measures (RCI and NI).

6.1 Different types of error

6.1.1 Errors in classification

SCANNER data points matched to ITN road sections having the wrong road classification may create errors due to the subsequent incorrect calculation of the RCI. As RCI calculations rely on knowledge of the road classification, matching points to a road section with a different classification to that upon which the data was actually collected can have a detrimental impact. For example, SCANNER data obtained on a C road which is fitted to a principal road will result in a much higher RCI value (if that data point shows bad condition) as a result of differing thresholds. As reported in section 5.1, a manual assessment of the fitting process showed that only 0.02% of the test dataset was matched to roads of a different class. This error rate is unlikely to cause any noticeable change in RCI and/or NI results at either a local or national level.

6.1.2 Errors in environment

Points matched to the wrong environment (urban/rural) may result in errors due to the subsequent incorrect calculation of the RCI. As discussed above (section 4.1) we proposed that the national system of classification be used to define an ‘Urban’ shape file. This would allow TOIDs within the ITN to be identified and assigned an urban or rural flag. This flag would be assigned at the section level, since this is the level of accuracy that is used in the UKPMS system. There is no need to assign the flag separately at the level of individual points. Applying this to the West Sussex road network gives the results shown in Figure 8.

![Figure 8: Rural and urban classifications in the West Sussex test data set. Red indicates urban roads, green indicates rural roads.](image)
The error that could be introduced by using this approach will arise from differences between the local authority’s definition of the environment and that defined within the national system of classification. An investigation into the agreement between the Local Authority (LA) and DfT definitions was therefore carried out for the West Sussex area. The results (Table 2) show the level of agreement both by section and by length of network. There are differences in the number of sections, with the DfT definition seeming to classify fewer sections as urban. However when considering lengths, the differences do not appear as significant. The Local Authority definition reports 29.6% of roads to be Urban and 70.4% to be Rural. The DfT definition gives 31.6% Urban and 68.4% Rural, thus 2% of the length of the network has been classified differently.

<table>
<thead>
<tr>
<th>Sections</th>
<th>Length (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LA</td>
</tr>
<tr>
<td>Urban</td>
<td>3732</td>
</tr>
<tr>
<td>Rural</td>
<td>217</td>
</tr>
<tr>
<td>Total</td>
<td>3949</td>
</tr>
</tbody>
</table>

Table 2: Rural and Urban classifications of roads on the UK network

The effect of a section being classified as rural is to slightly reduce the RCI. This has a knock-on effect on the local and national figures. The maximum change in RCI achieved by misclassifying a length would be 41\(^5\). This could promote a length within the ‘orange’ band of RCI values (20 – 100, ‘Plan Investigation Soon’) to the ‘green’ (generally good’, 0-20) band. Likewise a length in the ‘red’ or ‘Plan maintenance soon’ (100+) band could be promoted into the amber band. However, because of the calculation of the RCI such changes would only happen in cases where the texture depth (the only parameter within the RCI affected by Urban/Rural status) is of a low enough value to cause concern. The probability of this has been investigated using the West Sussex data set.

Within the West Sussex data, approximately 32% of the SCANNER measurements of texture depth had a value such that they would contribute to the RCI. As stated above, there was a difference of approximately 2% between the classification of road length allocated to the urban or rural environment by the local and DfT definitions. Therefore less than 0.7% of the RCI values from West Sussex would be affected by the combined influence of difference in environment and measured texture depth. Furthermore the effect on the RCI could be either positive or negative, so the allocation of different road environment will not always result in an increase in the RCI. Also, it is likely that a much smaller proportion of the network would be affected to the extent that the length would be changed from, say, amber to red and hence contribute to the NI. Therefore it is likely that the difference in environment will have only a small effect on the results generated from the fitted database.

\(^5\) The average change seen was 22.
6.2 Comparison of RCI values from UKPMS and RCI fitted SCANNER data

Note that direct numerical comparison of RCI values on a point for point basis is not straightforward due to the loss of section and chainage criteria for the data points, i.e. we cannot compare RCI values reported along the length of a road as a simple line graph and assess the numerical differences (this is discussed in section 4.4). Therefore, comparisons between the RCI values calculated using the original SCANNER data and the ITN fitted SCANNER data are most simply undertaken visually using a GIS. Here we may compare the RCIs using colour plots showing threshold RCI values represented as red, amber and green lengths.

Figure 9 and Figure 10 show the RCI values obtained over the complete West Sussex network. Figure 9 shows the RCI data calculated using a UKPMS system. Here the SCANNER data has been loaded into UKPMS, fitted to the West Sussex local authority network and RCI values calculated. Figure 10 shows the RCI data calculated using the processes developed in this work. Here the SCANNER data has been fitted to the ITN and RCI values calculated. Although it is not possible to make detailed comparisons at this scale, it can be seen that the overall results look very similar. In fact some of the differences are due to the slightly different ways the data is displayed, as lines from the UKPMS system and as dots from the fitted data.

![Figure 9: RCI data from UKPMS system](image-url)
A more detailed view of part of the data from Figure 9 and Figure 10 is shown in Figure 11. Here the red, amber and green dots show the RCI values obtained from the ITN fitted data (and are all placed along the centre line of the road coinciding with the ITN definition of the road). The lines drawn next to the road show the UKPMS representation, which are red, amber and green as before. Visual comparison between both data sets is very good. There are some offsets in the positioning of the red, amber and green lengths. This is as would be expected given the different methods of plotting the position of the readings. Hence, although the UKPMS data is shown here on a map it is still linked to the original section and chainage reporting by the SCANNER machine and hence is subject to the errors arising from the correct recording of section change points within the survey. However, the ITN data is being plotted using the OSGRs reported by SCANNER, following matching to the closest ITN road section. A particular example of the effect of this different approach is shown towards the bottom left of the picture, where the red dot (fitted data) is a distance from the red line (UKPMS data). This is most probably due to the UKPMS data being stretched to match section lengths (this is part of the process), whereas the fitted data is shown exactly where it was collected by the survey vehicle (following matching to the closest ITN road section).

Note that, at the local level errors in the data fitting process can affect RCI calculations by providing data at the wrong location. Data from one road can be fitted to another and thus result in an incorrect RCI. However, at larger scales such errors create less of an impact upon, for example, NI calculations. Data points cannot be fitted further than a pre-defined distance from their initial location. Therefore it is unlikely that a significant quantity of data will be fitted to a location outside the local authority from which the data was actually collected. Therefore, assuming that data points are typically fitted to the correct road type, NI calculations will not be significantly affected by errors in the fitting process. The calculation of NI values is discussed in the next section.
6.3 **Comparison of NI results from UKPMS and ITN fitted SCANNER data.**

The calculation of NI values can be used to provide an indication of the effect of error in the fitting process on both RCI values and the extent to which this would affect network level reporting. The values reported below in Table 3 are for West Sussex and in Table 4 for Birmingham. The data for West Sussex is based on only one year of survey data and that for Birmingham based on two year's of data. In the case of UKPMS results, the data has been selected and processed using a UKPMS comparable system. For the ITN fitted data, the data has been selected and processed using the rules and processes described in this document. Note that the NI values reported herein should only be used for assessment of the fitting process, and should not be compared with any published NI value for either authority. This is because we have made use of older SCANNER data, yet processed this using the latest rules for producing RCI.

For these calculations, we have used the method described in section 4 to identify C roads – that is by assuming that TOIDs from the ITN which have SCANNER data recorded against them and are not classified as A or B roads are in fact C roads. It should also be noted that the calculation for the NI169 from the fitted data has been based on the C road network length reported within UKPMS, due to the lack of data on C roads in the ITN. For the final version of the national database, data will need to be obtained or estimated for use in these calculations.
<table>
<thead>
<tr>
<th></th>
<th>UKPMS</th>
<th>ITN Fitted</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Selected Feature Lengths</strong></td>
<td>581.104</td>
<td>618.306</td>
</tr>
<tr>
<td><strong>Selected Lane Length</strong></td>
<td>968.541</td>
<td>997.771</td>
</tr>
<tr>
<td><strong>Surveyed Lengths</strong></td>
<td>544.554</td>
<td>547.24</td>
</tr>
<tr>
<td>% Selected</td>
<td>85.5</td>
<td>88.5</td>
</tr>
<tr>
<td><strong>Surveyed Lane Length</strong></td>
<td>923.594</td>
<td>922.22</td>
</tr>
<tr>
<td>% Surveyed</td>
<td>93.7</td>
<td>92.4</td>
</tr>
<tr>
<td><strong>Green &lt;40</strong></td>
<td>761.778</td>
<td>757</td>
</tr>
<tr>
<td>% Surveyed</td>
<td>82.5</td>
<td>82.1</td>
</tr>
<tr>
<td><strong>Amber &lt;=40, &lt;100</strong></td>
<td>133.271</td>
<td>135.31</td>
</tr>
<tr>
<td>% Surveyed</td>
<td>14.4</td>
<td>14.7</td>
</tr>
<tr>
<td><strong>Red &gt;=100</strong></td>
<td>28.546</td>
<td>30.19</td>
</tr>
<tr>
<td>% Surveyed</td>
<td>3.1</td>
<td>3.3</td>
</tr>
<tr>
<td><strong>NI Threshold Exceeded (&gt;100)</strong></td>
<td>27.30</td>
<td>28.94</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>B Roads</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Selected Feature Lengths</strong></td>
<td>329.44</td>
</tr>
<tr>
<td><strong>Selected Lane Length</strong></td>
<td>647.258</td>
</tr>
<tr>
<td><strong>Surveyed Lengths</strong></td>
<td>315.235</td>
</tr>
<tr>
<td>% Selected</td>
<td>95.7</td>
</tr>
<tr>
<td><strong>Surveyed Lane Length</strong></td>
<td>623.541</td>
</tr>
<tr>
<td>% Surveyed</td>
<td>96.3</td>
</tr>
<tr>
<td><strong>Green &lt;40</strong></td>
<td>360.505</td>
</tr>
<tr>
<td>% Surveyed</td>
<td>57.8</td>
</tr>
<tr>
<td><strong>Amber &lt;=40, &lt;100</strong></td>
<td>191.048</td>
</tr>
<tr>
<td>% Surveyed</td>
<td>30.6</td>
</tr>
<tr>
<td><strong>Red &gt;=100</strong></td>
<td>71.988</td>
</tr>
<tr>
<td>% Surveyed</td>
<td>11.5</td>
</tr>
<tr>
<td><strong>NI Threshold Exceeded (&gt;100)</strong></td>
<td>67.88</td>
</tr>
<tr>
<td></td>
<td>11</td>
</tr>
</tbody>
</table>

Table 3: ‘NI’ type results for West Sussex using UKPMS and fitted data.
### C Roads

<table>
<thead>
<tr>
<th></th>
<th>UKPMS</th>
<th>ITN Fitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selected Feature Lengths</td>
<td>1004.019</td>
<td>1004.019</td>
</tr>
<tr>
<td>Selected Lane Length</td>
<td>1968.872</td>
<td>1968.872</td>
</tr>
<tr>
<td>Surveyed Lengths</td>
<td>570.591</td>
<td>561.56</td>
</tr>
<tr>
<td>Surveyed Lane Length</td>
<td>569.249</td>
<td>561.56</td>
</tr>
<tr>
<td>Green &lt;40</td>
<td>285.25</td>
<td>279.42</td>
</tr>
<tr>
<td>Amber &lt;=40, &lt;100</td>
<td>180.132</td>
<td>178.77</td>
</tr>
<tr>
<td>Red &gt;=100</td>
<td>103.867</td>
<td>103.37</td>
</tr>
<tr>
<td>NI Threshold Exceeded (&gt;100)</td>
<td>98.20</td>
<td>97.78</td>
</tr>
</tbody>
</table>

### NI Results

<table>
<thead>
<tr>
<th></th>
<th>UKPMS</th>
<th>ITN Fitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>NI168 - A Roads</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>NI169 - Non-Principal Roads</td>
<td>16</td>
<td>16</td>
</tr>
</tbody>
</table>

---

**Table 3 (continued): ‘NI’ type results for West Sussex using UKPMS and fitted data.**

The results for West Sussex give an indication of the effectiveness of the fitting procedure for a large, mainly rural authority. Whilst there were differences in the methods used to define urban and rural (UKPMS using the local authority definition and the fitted data using the DfT definition), these do not seem to have influenced the results. Indeed, the 2% differences in urban and rural classification applied for West Sussex (between local definitions and national definitions) discussed in section 6.1 do not seem to have influenced the results, reflecting the fact that to affect the RCI requires the influence of other factors, such as texture depth.

The results for Birmingham show that, in a typical urban authority where the possibility for mismatching points is likely to be higher, the figures obtained are comparable when using a standard UKPMS approach and when using the fitted database.

As can be seen from Table 3 and Table 4, the NI results obtained using both methods (which are reported to the nearest integer) were identical. The figures for Green, Amber and Red were also comparable, with at most a difference of less than two for the percentage figures.

**Note:** Because of the assumptions necessary for the calculation of NIs in this report it should not be assumed that NI values presented herein are representative of the values that would be obtained by the named local authorities when reporting their performance indicators.
### A Roads

<table>
<thead>
<tr>
<th></th>
<th>UKPMS</th>
<th>ITN Fitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selected Feature Lengths</td>
<td>327.336</td>
<td>340.97</td>
</tr>
<tr>
<td>Selected Lane Length</td>
<td>424.771</td>
<td>438.36</td>
</tr>
<tr>
<td>Surveyed Lengths</td>
<td>251.144</td>
<td>244.03</td>
</tr>
<tr>
<td>Surveyed Lane Length</td>
<td>324.749</td>
<td>310.31</td>
</tr>
<tr>
<td>Green &lt;40</td>
<td>205.445</td>
<td>193.70</td>
</tr>
<tr>
<td>Amber &lt;=40, &lt;100</td>
<td>95.472</td>
<td>93.29</td>
</tr>
<tr>
<td>Red &gt;=100</td>
<td>23.832</td>
<td>23.32</td>
</tr>
<tr>
<td>NI Threshold Exceeded (&gt;100)</td>
<td>23.53</td>
<td>23</td>
</tr>
</tbody>
</table>

### B Roads

<table>
<thead>
<tr>
<th></th>
<th>UKPMS</th>
<th>ITN Fitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selected Feature Lengths</td>
<td>149.745</td>
<td>143.56</td>
</tr>
<tr>
<td>Selected Lane Length</td>
<td>236.042</td>
<td>235.36</td>
</tr>
<tr>
<td>Surveyed Lengths</td>
<td>144.801</td>
<td>143.56</td>
</tr>
<tr>
<td>Surveyed Lane Length</td>
<td>231.186</td>
<td>230.56</td>
</tr>
<tr>
<td>Green &lt;40</td>
<td>136.246</td>
<td>139.56</td>
</tr>
<tr>
<td>Amber &lt;=40, &lt;100</td>
<td>78.795</td>
<td>75.17</td>
</tr>
<tr>
<td>Red &gt;=100</td>
<td>16.144</td>
<td>15.83</td>
</tr>
<tr>
<td>NI Threshold Exceeded (&gt;100)</td>
<td>15.89</td>
<td>15.57</td>
</tr>
</tbody>
</table>

**Table 4:** 'NI' type results for Birmingham using UKPMS and fitted data.
### C Roads

<table>
<thead>
<tr>
<th></th>
<th>UKPMS</th>
<th>ITN Fitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selected Feature Lengths</td>
<td>343.6</td>
<td>343.6</td>
</tr>
<tr>
<td>Selected Lane Length</td>
<td>687.2</td>
<td>687.2</td>
</tr>
<tr>
<td>Surveyed Lengths</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Selected</td>
<td>27.4</td>
<td>27.5</td>
</tr>
<tr>
<td>Surveyed Lane Length</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Selected</td>
<td>13.6</td>
<td>13.5</td>
</tr>
<tr>
<td>Green &lt;40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Surveyed</td>
<td>74.4</td>
<td>74.3</td>
</tr>
<tr>
<td>Amber &lt;=40, &lt;100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Surveyed</td>
<td>21.0</td>
<td>21.1</td>
</tr>
<tr>
<td>Red &gt;=100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Surveyed</td>
<td>4.6</td>
<td>4.6</td>
</tr>
</tbody>
</table>

NI Threshold Exceeded (>100)

<table>
<thead>
<tr>
<th></th>
<th>UKPMS</th>
<th>ITN Fitted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

### NI Results

<table>
<thead>
<tr>
<th></th>
<th>UKPMS</th>
<th>ITN Fitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>NI168 - A Roads</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>NI169 - Non-Principal Roads</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 4 (continued): ‘NI’ type results for Birmingham using UKPMS and fitted data.
7 Implementation

7.1 Overall process for delivering a national SCANNER database

The fitting process described in section 3 constitutes one part of an overall methodology for processing SCANNER surveys and generating a national SCANNER database. The objective of this report has been to present the technical requirements for the fitting process, to support the delivery of a requirements specification for a production level system for the fitting of SCANNER data to the national framework. However, these processes will be required to fit within the overall process of delivering the SCANNER database. Table 5 suggests how the methodology described in this report may be implemented within a larger system.

7.2 IT requirements

7.2.1 Fitting and processing

It is suggested that the processes required to perform the implementation of the national SCANNER database be subsumed into one or more executable programs. These programs would undertake the following tasks:

- An initial pre-processing tool, which would require the user to select the suitable HMDIF files, and would then pre-process this raw data into the three tables discussed in stage 3 in Table 5, delivered in a suitable format for subsequent fitting.
- A second tool would perform the fitting process described in stage 4 in Table 5. The tool would also apply urban and rural labels to the data, according to stage 5 in Table 5.
- A further tool would generate the road condition measures according to stage 6 in Table 5. Note that the RCI for a particular point is calculated by applying the appropriate UKPMS weighting set to the observations recorded. Weighting sets are updated from time to time, so allowance needs to be made for new weighting sets to be used. Weighting sets may be provided in the form of an Access database, linking to this would be the easiest method of ensuring simple updating in the future.

A standard GIS interface would be used to allow a user to interact with the fitted data in an accessible visual format. RCI plots could be displayed in the GIS. The GIS would also provide reporting facilities, and the ability to calculate BVPI.

7.2.2 IT overhead

During development of the fitting processes a test algorithm was coded in Visual C++. This tool used pre-processed SCANNER HMDIF files as inputs and fitted them to an ITN network shape file. Processing was undertaken on a standard business PC running Windows XP. Processing of the Birmingham test data (to fit to the ITN) took approximately 30 seconds, the West Sussex data set took approximately 45 seconds. It is therefore felt that the national SCANNER database should not require IT systems having facilities beyond those found in a typical office.
The SCANNER data for one survey year is provided to DfT by the SCANNER survey contractors as a set of HMDIF files.

The HMDIF files are ‘pre-processed’ to extract the data in the format required for the application of the fitting algorithms. It would be practical for this stage to combine the SCANNER data into a single set of “tables” (see below) for subsequent use in the fitting process. It is suggested that further information be added to the SCANNER data at this point, such as the survey year.

Note 1: Later processing stages require knowledge of the road classification. It is not essential that this be provided because it can be determined following the completion of the fitting process. However, the ITN does not identify C and unclassified roads. It may be beneficial to the fitting process if the road class was provided via the SCANNER data. It would not be impractical for the SCANNER survey contractors to provide this information (for example by delivering the SCANNER HMDIF files on the basis of road classification). See section 4.1 for a discussion of this.

Note 2: The pre-processing system could be contained within a simple software standalone application, for speed of operation. See section 7.2.

The pre-processing produces three data tables, which are referenced to each other:

- Table 1: Containing sections from the SCANNER survey file (provides additional road data but is not used directly in the fitting process)
- Table 2: Containing points – the sets of OSGRs from the survey file. These will actually be fitted to the network.
- Table 3: Containing observations – the values recorded by the SCANNER survey.

The points data table is processed using the rules described in section 3 to fit to the pre-processed ITN network, which comprises

- Indexing
- Identifying the sub-set of ITN segments for fitting
- Comparison and scoring
- Application of post-processing

The resulting data set contains the original points, translated to appear on the ITN section to which they have been fitted. Because the tables are referenced to each other, the observations are also now referenced to the ITN.

Using the DfT supplied shape file for Urban/Rural within a GIS system, an additional field is added to the points table, and populated by selecting all points that lie within the urban area and assigning them a value of ‘urban’, with the remaining points being assigned a value of ‘rural’.

The observations data is processed to calculate RCI and BVPI – see section 4.

Table 5: Proposed stages in the fitting and processing process
7.2.3 Data storage

Estimates of the lengths of roads of different classifications on the English road network were provided in the specification for this project and are repeated here – see Table 6. The estimates are based on one line of data per 10m interval, although there may be more if the data are stored differently. In the sample data used for developing the fitting rules the locations of the data points were separated from their related condition data. For ease of processing, each observation was recorded on a separate row. In a final development of the database, all of the observations should be combined into one row per point.

<table>
<thead>
<tr>
<th>Road type</th>
<th>Length</th>
<th>Directions surveyed</th>
<th>Annual length surveyed</th>
<th>Estimated data lines for England annually</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>31,000 km</td>
<td>Two</td>
<td>31,000 km</td>
<td>3.1 million</td>
</tr>
<tr>
<td>B</td>
<td>20,000 km</td>
<td>Two</td>
<td>20,000 km</td>
<td>2 million</td>
</tr>
<tr>
<td>C</td>
<td>64,000 km</td>
<td>One</td>
<td>32,000 km</td>
<td>3.2 million</td>
</tr>
<tr>
<td>All classified local roads</td>
<td>115,000 km</td>
<td></td>
<td>83,000 km</td>
<td>8.3 million</td>
</tr>
<tr>
<td>Unclassified</td>
<td>179,000 km</td>
<td>Not yet covered</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Approximate distribution of road types in England

In the West Sussex test dataset the proportions of A, B and C roads reasonably approximate to the national figures, therefore it can be considered representative. One year of data for West Sussex occupies 240MB when stored in an Access database. Scaling this for the whole of England gives a figure of around 13.8GB of data per year. Allowing for the fact that observations (of which there are 6) were stored individually, this could be reduced by a factor of around 5 to give around 2.75GB of data. Whilst this is undoubtedly a considerable quantity of data, it should not represent an unreasonable storage proposition – 10 years worth of data would equate to less than 30GB of data, easily accommodated by current office IT (PC) hardware systems.

The choice of software database to hold the data may need to be considered as this quantity is too much for ‘desktop’ systems like Access (maximum capacity 2GB). Larger databases should easily be able to store a dataset of this size. MySQL can handle databases of “up to 2TB, possibly larger”\(^6\). Other larger scale databases could be expected to have similar capacities. Based on previous estimates, this should provide enough space to store over 600 years worth of data.

\(\text{\footnotesize \textsuperscript{6} http://dev.mysql.com/doc/refman/4.1/en/full-table.html}\)}
8 Conclusions and recommendations

To produce a practical central database that would allow the DfT to analyse and interrogate the data collected by the nationwide SCANNER survey (the national SCANNER database) it will be necessary to:

- Establish a single referencing system, or national framework, against which the data can be reported
- Develop appropriate rules to fit the SCANNER data to this framework
- Develop methods to process the data following fitting of the SCANNER data
- Develop methods for reporting the data from the central database

This project has carried out work to demonstrate how each of these requirements could be delivered.

Only a few networks exist that could be used as a national framework for referencing the SCANNER data. Of these, the Ordnance Survey MasterMap ITN is recommended as the most appropriate network for use in a national SCANNER database.

Processes have been developed for fitting the SCANNER data to the national framework. The fitting is carried out in three stages. Initial pre-processing of the delivered SCANNER data converts the data into a format suitable for fitting. This pre-processing extracts the location data from the SCANNER survey data for subsequent fitting to the network. Fitting algorithms are then applied to the location data to identify the appropriate Ordnance Survey ITN section to which the SCANNER data should be fitted. This process delivers the SCANNER data, but now referenced to the ITN. In this project the fitting process has been implemented in a standalone system as an “executable product”. It has been found that this delivers a high level of efficiency for the fitting process. However, the algorithms could be implemented in a conventional database if required. This could result in a significant increase in the processing time required for the fitting process.

Tests of the fitting have shown that the level of performance is high. Manual assessment of a random sample of data from a test network identified less than 0.5% of mismatched points, with less than 0.05% matched to roads of the wrong classification. Less than (approximately) 1% of SCANNER points were identified as “unmatchable”.

One area where the performance of the algorithms could be improved is in the fitting of SCANNER data collected on dual carriageway roads, where the data may be fitted to the wrong side of the dual carriageway. Algorithms have been suggested that may assist in improving the fitting of SCANNER data collected on dual carriageways. However, it is not recommended that these be implemented without further development. Furthermore, assessments using an urban test dataset showed that less than 0.1% of the data had been incorrectly fitted as a result of fitting to the wrong side of a dual carriageway.

Processing of the fitted SCANNER data can be carried out using the functionality provided by standard GIS and database systems. However, a number of external data sources are required in order to obtain measures such as the SCANNER RCI and National Indicators (NI). Data on the environment (urban/rural) can be provided from the national definition without significant effect on the results of the processing. Information on the classification can be obtained from the national framework. However, the recommended national framework (ITN) does not provide the facility to distinguish between C and U roads, which could lead to errors where local authorities have commissioned SCANNER surveys on these roads. It is suggested that SCANNER contractors be asked to provide the classification of the roads surveyed when providing the SCANNER data (for example by providing data from A, B, C and U roads in separate data files).
Following the completion of processing it is straightforward to apply standard rules for the calculation of National Indicators. The project has undertaken this calculation for two test networks, one primarily urban, the other primarily rural. The values of National Indicators NI168 (A roads) and NI169 (non-principal classified roads) were calculated for these test networks using both the standard UKPMS approach and using the fitted SCANNER database. Similar values were obtained for the indicators for both methods.

An implementation route for the database has been proposed, which recommends the use of dedicated software tools for the pre-processing and fitting of the SCANNER data, followed by the use of database and GIS facilities for the processing, display and reporting of the data. It is felt that the system could easily be implemented using typical office IT equipment. At the commencement of this project it had been suggested that there would be a need to process the SCANNER data from the delivered 10m lengths into 100m lengths in order to reduce storage space and optimise the processing time. However, the project has found that this should not be necessary.

**Acknowledgements**

The work described in this report was carried out in the Technology Development and Asset Management teams of the Transport Research Laboratory. The authors are grateful to Barry Cleave who carried out the technical review and auditing of this report.
Appendix A  Establishing the national framework

A.1 Introduction and requirement

TRL have been commissioned by DfT to produce a methodology for the fitting of SCANNER data to a national reference framework (road network) to allow data to be viewed and analysed on a national (rather than local) level.

In order to produce a national data set, there is a need for a system of referencing to be adopted for all of the data. SCANNER data is collected, processed and delivered with reference to Local Authority defined networks, and there is no national standard for the referencing (i.e. labelling of sections) of such networks. The first stage in the process is therefore to determine a suitable network to which to fit all of the SCANNER data.

SCANNER data consists of condition measurements, such as rutting and cracking, recorded during the survey and processed to provide summary data reported over 10m longitudinal lengths. In SCANNER network referencing is therefore still based on the traditional approach of section and chainage. However, during the SCANNER survey the equipment also records the geographical location of the survey vehicle (using GPS), which is again processed to provide co-ordinates reported at 10m intervals, broadly aligned with the condition data. Therefore, there is potential to fit the SCANNER condition data to an "external" network that has a geographical basis, using the co-ordinates provided within the SCANNER dataset. A logical approach to undertaking this fitting is to map the geographically defined scanner measurements (i.e. the known easting and northing for each condition measurement) to the required national network. To achieve this mapping in practice the network to which we wish to map the SCANNER data must be available in a ‘vector’ format.

There are few available sources of such a network. Indeed, the majority of datasets are based either partly or wholly on the Ordnance Survey MasterMap ITN (Integrated Transport Network). However, in addition to Ordnance Survey, TeleAtlas and Navteq also hold similar networks. These two companies provide the data used by almost all of the satellite navigation systems currently available.

We have therefore undertaken a brief review of these networks to determine which would be the most appropriate for use as the basis of a national SCANNER database. Note that the process of comparing the available networks was limited to three sources, as there would be no additional value in assessing the numerous networks available that are derived from these three source networks.

A.2 Comparison of networks

Table 7 summarises the major features of each dataset: Ordnance Survey ITN, TeleAtlas and Navteq.

There are several areas where the networks have common features, such as the ability to provide information in a vector format (deemed essential for practical fitting and reporting), and the provision of connected network lengths, with basic information on the road type. Road type data is probably essential. For example SCANNER data is not collected on roundabouts, but the errors in the recording of events during the survey could lead to data being geographically located on roundabouts. Knowledge of the road type could therefore assist in achieving a “clean” dataset. Similarly, because local authority location referencing of dual carriageway sections will differ from that used on two-way roads, knowledge of the road type on which the data was collected could be used to assist in the fitting of SCANNER data (e.g. when provided in both directions over a two year period).
Interestingly, none of the datasets is able to provide information other than location (e.g. provision of the road number) at a classification below B roads. There are no obvious additional fields that would assist in the identification of C roads in any data set.

For the current anticipated usage of the national SCANNER database this information is not considered essential, because we do not envisage that reporting of network data at the level of individual C road lengths (e.g. the “BVPI for the C123”) will be required.

The accuracy of the data contained within the networks appears to be high for all three sources. All of the companies have a process for regular updates to the network, with TeleAtlas and Navteq having vehicles that gather data for new and changed roads. Ordnance Survey on the other hand base their updates on the results of mapping exercises carried out to also update their bitmap and topological mapping. The update of the Ordnance Survey dataset is, perhaps, the most frequent of the three, and it will be more straightforward to monitor changes to the network over time, as this information is stored within the dataset.

All of the datasets provide additional information, beyond the basic vector mapping. Because both the Navteq and TeleAtlas networks have been collected primarily for use in route guidance systems the additional data stored alongside the basic shape of the network tends to concentrate on ‘points of interest’ for the driver, such as garages, parking, hotels etc. This additional information is probably of little interest for the national SCANNER database.
<table>
<thead>
<tr>
<th>Features List</th>
<th>OS ITN</th>
<th>Navteq</th>
<th>TeleAtlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vector Format network</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Connectivity</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Information on road type</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>(Single, Dual, Roundabout)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on road number</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Motorways</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A &amp; B roads</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>C roads</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Update Frequency</td>
<td>6 weekly</td>
<td>Not stated</td>
<td>Motorways and large cities are checked every year and secondary roads every two years.</td>
</tr>
<tr>
<td>Date of change</td>
<td>✓</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Additional</td>
<td>MasterMap Topology Layer - Gives accurate polygon representation of the roads, which may allow for greater ease and/or accuracy of fitting data</td>
<td>Data for additional features such as garages, car parks etc</td>
<td>Data for additional features such as garages, car parks etc</td>
</tr>
</tbody>
</table>

**Table 7**: Major features contained within each network
However, the potential additional information that could be provided within the Ordnance Survey dataset may be of use in the proposed application. The Ordnance Survey data has the ability to be linked with another dataset produced by OS – the Master Map Topographic Layer. This is another layer of the overall MasterMap product (of which the ITN is a part) that contains polygon representations of roads, tracks, paths, buildings etc. The difference between this data and that contained in the ITN is that there is an actual representation of the road, including width, giving the area covered by the road, rather than the ITN, which merely gives an indication of the approximate centre line of the road. It is felt that the addition of area could assist in the fitting process. It may be feasible to develop a fitting process where points within the SCANNER data (which will be reported against OSGRs) are selected based on the area into which they fall. This should assist at points where there is ambiguity as to the correct road to assign the data (i.e. at slip roads or where two roads run parallel and close to one another).

A.3 Conclusions and recommendations

The comparison of the three key networks available that could provide the basis for the national SCANNER database has not shown any particular network to have any remarkable strengths over another. Given this observation we must also consider additional factors when deciding on the network to use for the national SCANNER database. These lead us to recommend that the choice of network for use in mapping the SCANNER data should be the Ordnance Survey MasterMap, for the following reasons:

- The OS ITN contains as much data as any of the other datasets, but where there are differences these probably favour the use of the OS for the intended application. In particular the possibility of using the MasterMap Topology layer to assist in the fitting of the data may assist in areas where there are a high concentration of roads, and could improve the accuracy of the fitting process.

- The OS ITN is used throughout Government for the mapping and display of data. It would make sense to use the same referencing system for the National SCANNER Database to allow for easy exchange of information.

- DfT already have a license for the OS data. Whilst this may not be of relevance if there were a clear advantage to using one of the other datasets, given that they all are of similar suitability it would not be sensible to spend money to procure another dataset.

Given all of the above points, we recommend that the Ordnance Survey MasterMap ITN layer is used as the network for the mapping of SCANNER data.
Appendix B  The data fitting process

This appendix provides a more detailed description of the fitting process introduced in section 3. The fitting process comprises four major parts:

- Indexing
- Matching
- Scoring
- Further processing

The aim of the data fitting process is to take SCANNER data reported in relation to local authority sections and chainage and to reference this data in relation to ITN sections. This translation process draws on the OSGR data provided within the SCANNER survey data and ITN networks.

B.1 Input data

For the purpose of developing the fitting rules the location information in the SCANNER data is processed into a shape file and the following properties retained for use:

- SECID: The SCANNER section ID of the current section (normally an integer).
- POSID: The position ID of the SCANNER point within the current SCANNER section (normally an integer). Within each SCANNER section this starts at one and then increments with each point.
- XCO: The X-position of a SCANNER data point (OSGR Easting).
- YCO: The Y-position of a SCANNER data point (OSGR Northing).

The ITN consists of a shape file containing a series of roads. Roads bend and curve, hence each is represented by a series of small straight lines known as road sections. Each road section is typically of length 5 – 10m and has the following properties:

- DESCTERM, NATURE, CLASSIFICA, STREETNAME: These fields are used to generate a road description – useful in distinguishing between roads.
- TOID: The topographical ID of the road in the ITN. Each length of road has a unique TOID; within each TOID there are many road sections which describe the geometry.

Geographical (OSGR) location information is also included in the ITN shape file. This information is used in matching SCANNER data to the roads within the ITN.

These data form the input data to the indexing process.

B.2 Indexing

In order to match SCANNER data points to roads within the ITN, each point must be tested for compatibility against a number of ITN road sections. Ideally, the number of ITN road sections tested should be minimised, to increase the efficiency of the algorithm. This is achieved by indexing all the roads in the ITN into a three-dimensional array – where OSGR Eastings and Northings make up the first two dimensions and the direction of travel (bearing) is the third.

http://en.wikipedia.org/wiki/Shape_file
B.2.1 Indexing Longitude and Latitude

The ITN comprises a known geographical area covering the UK road network, that area can be split into an array of $n$ by $m$ by $z$ boxes – or array segments. Each ITN road section fits into one or more of these segments depending on its location and direction (each ITN road section is a straight line and can only have one direction).

The algorithm reduces the number of ITN road sections against which each SCANNER point must be tested by first matching each SCANNER point to an array segment, then testing the SCANNER point against only those ITN road sections found in that array segment and segments adjacent to it.

The principle of this is demonstrated in Figure 12. The SCANNER point (dark purple) is located in the dark green central array segment. Neighbouring array segments are light green. Potential ITN road sections for matching (see below) the SCANNER point include those inside the green array segments, i.e. those road ITN sections marked in red.

![Figure 12: A 2-dimensional array of segments for OSGR Eastings and Northing.](image)

The size of each array segment may be adjusted to allow fine-tuning of the accuracy of the algorithm. Using a large number of array segments (small array size) reduces the number of ITN road sections that each SCANNER point is compared with – thus increasing efficiency. Using a small number of array segments (large array size) increases the number of ITN road sections that each SCANNER point is compared with – increasing the chance that a SCANNER point will be matched and decreasing efficiency.

Generally when creating an index for an urban area with many roads, it is desirable to have a large number of array segments (high resolution) to cope with the large number of closely-packed roads. In a rural area a smaller number of array segments (low resolution) would be preferable, as the smaller number of roads would not compromise the accuracy of the algorithm.

For the purposes of developing the rule set, the size of each array segment was set to a constant size of 100m.
B.2.2 Indexing direction of travel

The direction of travel (bearing) is indexed in order to provide a third dimension on to which SCANNER points are fitted. In the fitting process the SCANNER point bearing is defined as an angle from 0° (northbound) to 360° (also northbound) where 90°, 180° and 270° are east, south and west respectively.

Figure 13 demonstrates how indexing on bearing improves efficiency on a fictitious junction where 2 dual-carriageway roads cross. The SCANNER points have been reported in the SCANNER data file in a direction of travel starting at the northeast and going southwest. After indexing on only OSGR Eastings and Northings, all ITN road sections shown in Figure 13 are in the same array segment. Therefore all road sections (red and grey) would be compared against the SCANNER points. After indexing on bearing, only the ITN road sections marked in red would be considered for matching with the SCANNER points, since only those ITN road sections have the same bearing.

![Figure 13: Indexing on bearing](image)

Notice that the ITN road sections for both carriageways are eligible to be compared with the SCANNER points, even though the bearing of traffic on one carriageway is opposite to the bearing on the other. This is because road bearings (as opposed to SCANNER point bearings) stored in the ITN run only between 0° and 180°. A north/south road has a bearing of 0° while an east/west road has a bearing of 90°. The opposing carriageways will have the same bearing in the ITN. Hence bearing information in the ITN is independent of the direction of flow. See Appendix C.

Similarly to indexing on position, a decision must be made as to the size of the index segments created. For the purposes of developing the rule set a segment size of 30° was used. It is advisable that a fairly large angle is used, since the directional accuracy of the SCANNER data is unknown.

B.3 Matching

Indexing the ITN enables the rapid reduction of potential ITN road sections against which each the SCANNER data point needs to be tested (for fitting). However, even after indexing to a shortlist of potential ITN road sections, there is still a requirement to assess the SCANNER data point against each ITN road section order to identify the
correct one. This is done by assessing the likelihood that the SCANNER point comes from the ITN road section.

Three calculations must be carried out in order to calculate the likelihood of a match against each road section; direction, projection and distance.

**B.3.1 Direction**

The direction in which the SCANNER vehicle was driving can be estimated by drawing a line from the previous SCANNER point, through the current one and the on to the next.

At the start of a SCANNER section there is no ‘previous’ point while at the end of a SCANNER section there is no ‘next’ point. This problem can be resolved by substituting the current SCANNER point for the missing one. However, this may introduce errors where there is a change of SCANNER section on a sharp bend.

**B.3.2 Projection**

After determining direction of the SCANNER data point, the position and the bearing of the point are known. The index array segment corresponding to this position and bearing is found and the SCANNER data point is tested against each ITN road section therein. The test involves projecting the point on to each ITN road section and finding the distance between the point and projection.

The SCANNER point is projected onto each potential ITN road section by finding the closest point to the SCANNER point which is on the ITN road section. If the road is perfectly horizontal or vertical then this is trivial. If not then the projected point \((X_a, Y_a)\) is given by:

\[
X_a = \frac{(X_p + m(Y_p - c))}{1 + m^2}
\]

And

\[
Y_a = mX_a + c
\]

where the SCANNER point is at \((X_p, Y_p)\) and the road section is a line with equation \(Y = mX + c\). Note the equation \(Y = mX + c\) represents a line stretching to infinity which contains the road section, this means that although the point \((X_a, Y_a)\) will always lie on the line \(Y = mX + c\) it may not lie on the ITN road section itself. When this occurs the projected point is moved to the end or start of the ITN road section – whichever is nearest.

**B.3.3 Distance**

The distance between the SCANNER point and the projected point is calculated.

**B.4 Scoring (likelihood)**

After a SCANNER point has been matched against every ITN road section in an index segment, the matches are scored to determine the best fit. The score is calculated by summing the distance between the projected and reported point with half the difference in bearing between the two points (in degrees). A higher score is considered to describe a poorer match, zero describes a perfect match – see Appendix D.

Although this scoring methodology is arbitrary, it appears prudent to base the scoring on the available information regarding the quality of match between SCANNER points and road index. The further a SCANNER point lies from an ITN road section and the greater difference between the respective bearings, the less likely a road section is as a candidate for matching the SCANNER point.
The ITN road section giving the lowest score is potentially selected as the matching section. The application of moving average smoothing to the scores has potential to alter this selection.

**B.5 Further processing**

At this stage in the fitting process each SCANNER data point has been located within an ITN array segment and scored against each ITN road section therein. Although at this stage it is possible to accept lowest (best) score as a match for each SCANNER point, there are additional considerations which may influence this decision.

**B.5.1 Should the closest match (lowest score) always be selected?**

Consider Figure 14, the situation depicted shows a number of SCANNER points (purple) which are being fitted to ITN road sections (red) using the fitting rules as described. The part of the ITN road network shown consists of a dual carriageway with an exit slip (green) on the northbound side. The SCANNER points have for the most part been fitted to the main carriageway. However, 2 SCANNER points are fitted to the slip road – see the scores presented in Table 8.

![Figure 14: Fitting SCANNER points at a slip road.](image)

<table>
<thead>
<tr>
<th></th>
<th>Slip Score</th>
<th>Northbound Score</th>
<th>Southbound Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Point 1</strong></td>
<td>50</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td><strong>Point 2</strong></td>
<td>50</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td><strong>Point 3</strong></td>
<td>35</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td><strong>Point 4</strong></td>
<td>20</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td><strong>Point 5</strong></td>
<td>10</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td><strong>Point 6</strong></td>
<td>10</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td><strong>Point 7</strong></td>
<td>20</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td><strong>Point 8</strong></td>
<td>35</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td><strong>Point 9</strong></td>
<td>50</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td><strong>Point 10</strong></td>
<td>50</td>
<td>20</td>
<td>40</td>
</tr>
</tbody>
</table>

*Table 8: Approximate scores obtained when matching SCANNER points against the ITN road sections shown in Figure 14.*
Although the fitting algorithm has succeeded in finding the closest ITN road section to each SCANNER point, a visual inspection of the data reveals it is wrong – since the SCANNER vehicle continues on the dual carriageway it cannot possibly have driven on the slip road.

Since the SCANNER vehicle can only transfer between roads at junctions, continuity in the location of SCANNER points is expected. Discontinuities in the locations of matches are likely to indicate that the matching algorithm has made a mistake.

**B.5.2 Moving average smoothing**

Consider the likelihood scores for the 10 SCANNER points closest to the slip road – given in Table 8. The northbound carriageway has consistently good scores. The southbound carriageway has consistently bad scores. For points 5 and 6 the slip road scores best, hence those two SCANNER points are matched to the slip rather than the main carriageway. A moving average which gives a higher weighting to scores close to the current SCANNER point can be used to smooth potential discontinuities in SCANNER point matched locations. This provides an alternative score, which results in all SCANNER points in Figure 14 being fitted to the northbound carriageway.

The moving average considers the average match scores of 11 points, the SCANNER point itself and the matches obtained on the same road by the closest 5 SCANNER points on either side of it. The weighted moving average $X_i$ is

$$X_i = x_i + \sum_{j=1}^{5} (0.75)^j (x_{i-j} + x_{i+j})$$

where $x_i$ are the match scores to a given road at SCANNER point $i$. This is applied to the scores in Table 8 (where the scores for point 1 and point 10 have been extended prior to and beyond the data set respectively) and results in scores shown in Table 9.

<table>
<thead>
<tr>
<th></th>
<th>Slip Score</th>
<th>Northbound Score</th>
<th>Southbound Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Point 1</strong></td>
<td>235.56</td>
<td>111.52</td>
<td>223.05</td>
</tr>
<tr>
<td><strong>Point 2</strong></td>
<td>214.03</td>
<td>111.52</td>
<td>223.05</td>
</tr>
<tr>
<td><strong>Point 3</strong></td>
<td>188.88</td>
<td>111.52</td>
<td>223.05</td>
</tr>
<tr>
<td><strong>Point 4</strong></td>
<td>167.65</td>
<td>111.52</td>
<td>223.05</td>
</tr>
<tr>
<td><strong>Point 5</strong></td>
<td>154.67</td>
<td>111.52</td>
<td>223.05</td>
</tr>
<tr>
<td><strong>Point 6</strong></td>
<td>154.67</td>
<td>111.52</td>
<td>223.05</td>
</tr>
<tr>
<td><strong>Point 7</strong></td>
<td>167.65</td>
<td>111.52</td>
<td>223.05</td>
</tr>
<tr>
<td><strong>Point 8</strong></td>
<td>188.88</td>
<td>111.52</td>
<td>223.05</td>
</tr>
<tr>
<td><strong>Point 9</strong></td>
<td>214.03</td>
<td>111.52</td>
<td>223.05</td>
</tr>
<tr>
<td><strong>Point 10</strong></td>
<td>235.56</td>
<td>111.52</td>
<td>223.05</td>
</tr>
</tbody>
</table>

**Table 9: Showing the results of applying a moving average to the scores in Table 8. Each road is considered separately when calculating the moving average.**

The scores for the northbound carriageway remain good. The scores for the southbound carriageway remain poor. Points 5 and 6 are no longer best matched to the slip road. The weighted average over adjacent points has filtered out the jump in scores seen in Table 8. The smoothing of the scores allows the northbound carriageway to be selected as the best match for all SCANNER points; the presence of the slip road no longer causes errors.
This approach allows the algorithm to determine where jumps in matching have occurred, and to compensate without preventing correct matching of SCANNER points when the vehicle has in fact driven onto a new road. If the SCANNER vehicle was to actually drive onto the slip road shown in Figure 14, enough good-scoring matches with the slip road would provide a good moving average score and allow the algorithm to match the SCANNER points correctly.
Appendix C  Dual carriageways

Consider the indexing of the dual carriageway in Figure 13. Notice that both the north and south bound carriageways have the same bearing even though the flow of traffic is in opposite directions. This arises because the ITN contains limited information on dual carriageway roads. Dual carriageways are identified as “independent” roads. However, no information about the direction of flow on the carriageway is provided.

To make good use of the reporting of SCANNER data it is desirable to be able to fit the SCANNER data to the correct carriageway on dual carriageway roads. Unfortunately the error in the SCANNER position measurement can lead to difficulty in deciding the carriageway to which a particular SCANNER point should be fitted. Clearly, a simple way to decide on the appropriate carriageway would be to calculate the direction in which the SCANNER survey vehicle was travelling (from the SCANNER data) and then select the carriageway having the same direction. Hence the ability of the algorithm to match SCANNER points to the correct side of a dual carriageway would be improved by aligning the direction of travel of the vehicle with the direction of flow of the road.

Information about the direction of flow can be inferred from the ITN network by matching ITN dual carriageway road sections against ‘partner’ ITN dual carriageways nearby. Two ITN road sections are considered a dual carriageway ‘pair’ if:

1. The *nature* of both road sections is ‘Dual Carriageway’
2. The *descriptions*\(^8\) of both road sections are identical.
3. The TOID of each road section is different.
4. The road sections share no common start or end nodes.
5. The road sections are fewer than 10m apart\(^9\).

Italicised terms in the above list refer to fields contained in the ITN network.

For each dual carriageway pair, a ‘left’ and a ‘right’ side (henceforth referred to as the LHS and RHS respectively) is determined by projecting one on to the other – thus forming a general impression of the directionality. See Figure 15.

Figure 15: The calculation of directionality is demonstrated for two dual carriageway pairs. The dotted red line shows the projection from the shortest to the longest of the pair.

---

\(^8\) The ‘description’ is defined as the concatenation of three fields in the ITN; *Descterm*, *Streetname* and *Classifica*.

\(^9\) 10 m was used during testing.
By always projecting from the shortest to the longest of the pair, a reference point is formed from which to assign directionality. If the projection is between 0° and 180° then the shortest of the pair is the LHS, if the projection is between 180° and 360° then the longest of the pair is the LHS. Using this system provides a rotational symmetry – provided that you do not rotate through more than 180°. This is demonstrated in Figure 16.

Figure 16: Demonstrating the rotational properties of the definitions of the left and right hand sides of the dual carriageway.

This works well for north / south roads, but is prone to failure on east / west roads. Since an east / west road travels along a bearing of approximately 90°, the projection from one side of the dual carriageway to the other will fluctuate around either 0° or 180°. Hence the RHS and LHS classifications will become confused – see Figure 17.

Figure 17: Showing a potential error when assigning left and right hand sides to roads. Consecutive road sections on a given TOID are – because of the road direction – assigned to the LHS and RHS respectively.

The problem identified in Figure 6 can be overcome by determining an overall directionality for each TOID in the ITN. This is done by simply summing the number of LHS and RHS road sections and using whichever is the greatest to provide the TOID directionality. After categorising all dual carriageway road sections as either LHS or RHS carriageways, the heading of each TOID in the ITN can be estimated by considering the road bearing and which side of the dual carriageway it is on – see Table 10.

<table>
<thead>
<tr>
<th>Bearing</th>
<th>LHS</th>
<th>RHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-45°, 135° - 225°, 315° - 360°</td>
<td>Northbound</td>
<td>Southbound</td>
</tr>
<tr>
<td>45° - 90°, 225° - 270°</td>
<td>Eastbound</td>
<td>Westbound</td>
</tr>
<tr>
<td>90° - 135°, 270° - 315°</td>
<td>Westbound</td>
<td>Eastbound</td>
</tr>
</tbody>
</table>

Table 10: The directionality of each section on a dual carriageway is determined by the road bearing and the side of the carriageway.

Figure 18 shows the results of applying this method to dual carriageways in the Birmingham ITN, showing the designation of dual carriageway segments as left hand side and right hand side. A region, circled and labelled ‘A’ in Figure 18, is enlarged in Figure 19. Although the method has been broadly successful, the region has been selected to demonstrate how the left hand and right hand carriageway designations can be less successful when the road runs in an east / west direction. A manual inspection of a randomly-selected set of data from the Birmingham area showed that approximately 87% of points were correctly matched to either a left or right hand carriageway.
road which travelled broadly north to south (the A34 between the A4041 and the A4040) this accuracy rose to 95%. Carrying out this analysis on a road which travelled broadly east to west (the A4182 in central Birmingham) showed a decrease in accuracy to 82%.

Figure 18: Showing the distribution of right and left hand points on dual carriageways in the Birmingham. Black points are right hand, yellow points left hand.

Figure 19: Left hand and right hand side designations become mixed where the dual carriageway runs east to west.

C.1 Using this method to improve fitting of SCANNER data on dual carriageways

The methodology described above enables the direction of flow to be estimated on ITN dual carriageways. Note that the calculation of the direction of flow only needs to be carried out once, for example as part of a pre-processing stage applied to the ITN network. The pre-processed ITN data would then contain information on the direction of
When fitting SCANNER data to a dual carriageway, the known direction of travel of the SCANNER vehicle (obtained from the SCANNER OSGR data, as described above in Section 4) would be compared with the estimated directionality of the ITN section (obtained in the pre-processing of the ITN data) to assist in selecting the correct ITN road section.

An analysis of this methodology revealed that only 30% of points were moved to the correct location – i.e. 70% of points moved ended up in the wrong place as a result.

It is recommended that this methodology not be included in the fitting algorithm in its present form for the following reasons:

- A review of the raw fitted SCANNER test data in Birmingham revealed that less than 1km of data contained points which had been fitted to the wrong side of a dual carriageway (< 0.1%).
- The number of errors created by the methodology is greater than the number of errors corrected.
- The problem of data points drifting may be less prevalent on other parts of the network. The proportion of dual carriageways on the national network is far less than in a large urban city such as Birmingham.

It is further recommended that additional work to investigate improvements in the methodology be carried out. Basing the decision to move a point from one side of a dual carriageway to the other on a more stringent set of criteria could reduce false positives at the expense of correcting some errors. In addition, future changes to the ITN may include additional information which would aid the identification of dual carriageways and further reduce the rate at which errors are created.
Appendix D  The scoring process to determine the likelihood of matching

During the matching process, each SCANNER data point is matched against the road section for which the best likelihood can be calculated. The likelihood of a match is calculated by adding the distance between the SCANNER data point and the projected point to half the difference in angle between the two.

Consider an example of SCANNER points being fitted to a road at a point where two roads cross at a right-angled intersection – see Figure 20.

![Figure 20: SCANNER data points on a NE / SW road. Distances from the roads are provided to aid in the likelihood calculation, bearings of the roads and SCANNER points are also provided](image)

Each point is potentially projected on to both the NE / SW road and the NW / SE road, both cases are considered and the details of each projection calculated. To illustrate this, a single point is highlighted and its scores determined – see Figure 21.

![Figure 21: The upper SCANNER point can be either projected to the NW / SE road (A) or the NE / SW road (B)](image)

In the example, projection A is a distance of only 1m while projection B is a distance of 5m. The bearing of the SCANNER points (calculated by drawing a line from the previous point to the next one) is 40°. The bearing of the road matched to in projection A is 135° while the same for projection B is 45°.
The likelihood scores for each projection are:

$$1 + \frac{135 - 40}{2} = 48.5 \text{ for projection A and}$$

$$5 + \frac{45 - 40}{2} = 7.5 \text{ for projection B.}$$

Projection B is the clear winner. Using this system of likelihood scoring ensures that a large proportion of SCANNER points near junctions and intersections will be matched to a road that is both nearby and also runs on the same bearing. Further post-processing ensures that even where SCANNER points are matched to the wrong road by selecting the lowest likelihood score, there is a good chance that they will be moved back into the right place – see Appendix B.5.
Development of network fitting and data handling procedures for a national SCANNER database

The SCANNER survey provides local authorities with an objective assessment of the condition of the (surveyed) local road network. Although there is considerable potential for the use of the SCANNER data both locally and nationally, centrally the DfT sees only the summary data reported as a National Indicator for each Local Authority. By storing the SCANNER data centrally the DfT would have access to a powerful database of network condition. Such a database would provide the potential to undertake a wide range of analyses. However, to provide such a central database there is a requirement to develop methods to take the data from the different local authority networks, surveys and survey contractors and fit this data to a single defined reference framework, against which each data set can be aligned, displayed and reported. There is then a need to develop procedures for the processing of the data so that additional information can be obtained (for example Road Condition Index (RCI) or NI values).

TRL was therefore commissioned by DfT to produce a methodology for the fitting of SCANNER data to a national reference framework (road network) to allow data to be viewed, processed and analysed on a national (rather than local) level. The results of this work are presented in this report.

Other titles from this subject area

PPR315 Measuring skid resistance without contact. A Dunford. 2008
PPR299 Automated detection of fretting on HRA surfaces. S McRobbie and G Furness. 2008
PPR253 Investigation of the effects of pavement stiffness on fuel consumption. E Benbow, J Iaquinta, R Lodge and A Wright. 2008