PUBLISHED PROJECT REPORT PPR238

SCANNER Condition Indicator parameter thresholds and weightings

Version: 1.0

by S McRobbie

Prepared for: Project Record: Developing SCANNER Condition Indicator parameter thresholds and weightings
Client: RLTSF(3), Department for Transport, (Edward Bunting)

Copyright TRL Limited March 2007

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

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

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

Executive summary ii

1 Introduction 1

2 Objectives 1

3 Methodology 2

3.1 Realisation of the Defects Index model 2
3.2 Comparison of model results and reference data 3
3.2.1 Reference Data 3
3.2.2 Model Results 3
3.2.3 Approach taken to comparison of data 4
3.3 Network level analysis 4

4 Model realisation 5

4.1 Input data 5
4.2 Model operation 5
4.3 Initial model settings 6

5 Data used 8

5.1 Comparison with engineers’ assessments 8
5.1.1 Distribution of site lengths in reference data 8
5.1.2 Distribution of site ratings in reference data 8
5.1.3 Occurrence of specific defects, and seriousness of defect in reference data 9
5.2 Network level assessments 10

6 Results and discussion 11

6.1 Comparison with engineers’ assessments 11
6.1.1 Results using default settings. 11
6.1.2 Effect of changing weightings 13
6.1.3 Effect of changing thresholds 14
6.1.4 Effect of non-linear relationship for weighting rut depth 15
6.2 Network level assessments 15

7 Recommendations 17

Acknowledgements 19

References 19

Appendix A. 20

Appendix B. 22
Executive summary

In order to provide an improved measure of the maintenance condition of the road network the Department for Transport (DfT) has introduced SCANNER surveys on local roads in England and Wales. These surveys are performed using vehicle based automatic monitoring equipment, and replace routine visual condition surveys such as CVI or DVI. The surveys monitor the condition of the roads by making measurements and determining the values of a range of parameters such as cracking, rutting, longitudinal profile and surface texture.

There is a need for a consistent analysis method for SCANNER survey results, such as a condition indicator or defects index. This will facilitate a reliable quantitative comparison between the condition of roads in different parts of the country, and enable the change in condition over time to be monitored.

The Defects Index Working Group (DIWG) was established from a group of Local Authority engineers and other interested parties. The DIWG was responsible for the delivery of a preliminary definition of a Defects Index (CBC, 2004). This preliminary definition identified which parameters should be included in the Defects Index, what the initial settings might be for the weightings and thresholds, and made some suggestions for possible variations to the model. The initial report also made some suggestions regarding the way the data should be reported, and on the number of maintenance categories to be reported.

TRL have taken the recommended Defects Index model as proposed by the DIWG and CBC and have used real data to test it. These tests have examined different methods for combining and aggregating the data, different threshold levels and weightings, and different ways of reporting the condition of a given 10m subsection of road.

This has been done by obtaining Engineering assessments of a number of sites from a number of Local Authorities. These opinions were then used as a reference data set in a series of comparisons between the output of the Defects Index model in its various forms.

Using crude statistical approaches, the work identified the parameters which showed the strongest relationship with the engineers’ opinions. These were found to be the 3m and 10m LPV parameters. It was also found that rutting was less closely linked to the engineers’ opinions of the sites than was expected.

The research also found no strong basis for adjusting the threshold levels used in the model. The thresholds used had been set based on extensive prior study and were set, in some cases, for safety reasons. It was felt that this study did not offer sufficient scope for varying these thresholds and overruling the earlier, more in depth work.

The use of ‘families’ of data, in which similar defects were grouped together and only the most serious defect in any given family was used to calculate the index, was investigated. It was felt that this approach definitely made sense for the nearside and offside rutting measurements, and so was adopted for these parameters.

A brief study was performed into the use of non-linear weighting curves of defects between the lower and upper thresholds. No evidence was seen that this was beneficial at this stage.

Various methods were tested for combining and reporting the Defects Index. These included the sum of all the weighted parameter values, the average of the weighted parameter values, a weighted average of the weighted parameter values and reporting only the most serious defect in each 10m subsection. Reporting the sum of the weighted parameter values was chosen as the best method.
1 Introduction

In order to provide an improved measure of the maintenance condition of the road network the Department for Transport (DfT) has introduced SCANNER surveys on local roads in England and Wales. These surveys are performed using vehicle based automatic monitoring equipment, and replace routine visual condition surveys such as CVI or DVI. The surveys monitor the condition of the roads by making measurements and determining the values of a range of parameters such as cracking, rutting, longitudinal profile and surface texture.

There is a need for a consistent analysis method for SCANNER survey results, such as a condition indicator or defects index. This will facilitate a reliable quantitative comparison between the condition of roads in different parts of the country, and enable the change in condition over time to be monitored.

The Defects Index Working Group (DIWG) was established from a group of Local Authority engineers and other interested parties. The DIWG was responsible for the delivery of a preliminary definition of a Defects Index, which was developed under the initial TTS research programme (CBC, 2004). This preliminary definition identified which parameters should be included in the Defects Index, what the initial settings would be for the weightings and thresholds, and made some suggestions for possible variations to the model. The initial report also made some suggestions regarding the way the data should be reported, and on the number of maintenance categories to be reported.

It was recommended that further work should be carried out prior to the implementation of the Defects Index. The recommended work included an assessment of the suitability of the thresholds and weightings proposed for use with the Defects Index on a sample local roads.

This further work has examined the settings proposed by the DIWG, following the initial research, and has tested some of the variations proposed.

2 Objectives

The primary objective of this project was to carry out the further testing required by the DIWG prior to the implementation of the Defects Index for use with SCANNER data.

This testing has taken place on a sample of local roads. Reference data as to the actual condition of the roads, for comparison with the Defects Index, has been provided in the shape of engineers’ assessments, carried out specifically for this work. The research has examined the suitability of the proposed settings for the thresholds and weightings to be used in the model.

The research has also briefly investigated the various options for how the data should be processed and amalgamated within the model in an attempt to find out whether any of the more complex methods proposed provide an advantage over the basic model. This has included an investigation of the effect of non-linear weighting curves, and of different ways of combining the weighted parameter scores to obtain a single value for each reporting length.
3 Methodology

A three stage process was undertaken to determine the best values, and the most appropriate methods of combination for the different defects, thresholds and weightings used in the model. The three stages were:

1. **Realisation of the Defects Index model, in an easily parameterisable and adjustable form.**

   This enabled the different weightings, thresholds and options for combining data to be easily tested. This step was necessary before any performance testing could take place.

2. **Comparison of model results and reference data on a number of specific sites.**

   By comparing the output of the model with the reference data, and how this output varied as parameters were varied, we were able to determine which model inputs were most important in obtaining a model result which agreed with the engineers’ assessments of the reference sites. We were also able to test different values for the weightings and thresholds and see which values seemed most appropriate.

3. **Network level analysis, performed by processing all available SCANNER data from a selection of Local Authorities.**

   By considering the proportions of the networks which were classed in each of the maintenance categories we were able to see whether or not these seemed appropriate.

3.1 **Realisation of the Defects Index model**

The realisation of the Defects Index model is discussed in more detail in Section 4. The process of model realisation involved converting the model from the specification and recommendations included in the initial report (CBC, 2004) into a computer model which would allow real data to be processed and produce output following the guidelines of the DIWG.

The realisation of the model took place in Microsoft Excel using basic Visual Basic macros. The model allowed the input data to be imported, formatted, weighted and combined in a variety of ways before results were produced.

The software involved in the model realisation included some basic checks on the quality and integrity of the data. The model was not designed to be an operational and fully robust piece of software and, as such, could not deal with all the possible issues which the data presented. Therefore, any unsuitable lengths were manually filtered out from the results.

The model was written in such a way that different options could be easily enabled or disabled and modifications could be easily implemented and tested. For example it was possible to use the model to determine the condition of a given stretch of road based on any one of several ways of combining the weighted parameter scores.

The inputs required by the model were the 10m values for the following parameters:

- Nearside and offside rut measurements;
- 3m and 10m Longitudinal Profile Variance (LPV) values;
- Nearside SMTD texture measurements;
- Whole carriageway crack values;
- Nearside and offside wheeltrack cracking values.

The model processed the data and assigned each 10m subsection within the data into one of three condition categories, referred to as ‘Green’, ‘Amber’ and ‘Red’. Green subsections were those which
were in sound condition and required no maintenance, Amber subsections were those which were not in need of urgent attention, but which exhibited some defects which may cause the engineer concern, and Red sites were those which had a need for maintenance attention quite soon.

3.2 Comparison of model results and reference data

3.2.1 Reference Data

In order to determine how well the model performed, TRL compared the output of the model with reference data consisting of engineering assessments of a number of sites. Local Authority engineers were asked to select a number of sites to be used in the assessment, and give an engineering assessment of the extent and type of deterioration for each.

The characteristics of suitable sites, and the number of sites required from each authority to make the analysis meaningful were considered and a document (Appendix A) was prepared. This document was sent out to a number of Local Authorities selected from the SCANNER IAG, with a few other interested parties. The document explained the type of sites suitable for inclusion, and what the engineer performing the site assessment was expected to consider and report. The constraints included site length, data availability, and the nature of any defects present on the site. An additional document (Appendix B) was also sent out showing the format in which the results were to be provided.

TRL proposed the following rating system for use when selecting and assessing the sites (Table 1).

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>Sound condition</td>
<td>Local deterioration</td>
<td>Partial deterioration</td>
<td>General deterioration</td>
<td>Urgent maintenance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 1: Proposed site rating scheme for selection of reference sites.*

Each Local Authority was asked to provide details of 4 sites which they felt were in sound condition, 4 sites exhibiting local or partial deterioration, and 4 sites showing general deterioration, or that were in need of urgent maintenance. Recently maintained sites were to be excluded from the assessments and sites were also to be limited to those exhibiting defects capable of being detected using SCANNER technology, such as cracking, rutting, roughness or certain texture defects.

The engineer performing the assessment was asked to note the type and rate the severity of the defects on each site. The proposed ratings for this assessment were:

- 0 – Not present
- 1 – Present
- 2 – Present and a problem
- 3 – Major problem

3.2.2 Model Results

SCANNER data from participating Local Authorities was provided in HMDIF format. The data from the reference sites was extracted and processed using TRL’s proprietary ChartCrack software in order to get it into a format suitable for input to the model. This processed data was then imported into the model, and the results of the model were compared to the engineers’ assessments.

The model settings could then be adjusted, and the data rerun through the model. The new results, obtained using the modified settings, were compared to the reference data, and the original results to see the effect of the altered settings.
This process was repeated several times until the final configuration of the model, and the optimum parameter thresholds and weightings were settled upon.

### 3.2.3 Approach taken to comparison of data

Two main methods were used for quantitatively assessing the performance of the model using any given set of weightings and threshold values. The first method was to calculate the proportion (by length) of sites which were classed by the model as being either Red or Amber, and to compare this against the engineers’ assessment of the site condition. The second approach was to calculate the average score for each site, based on the scores obtained for each of the 10m subsections within the site. These average scores were also compared to the engineers’ assessments.

The proportion of each site classed as Red or Amber gave an indication of the extent of the defect along the site, whilst the average score gave some indication of how severely the section had deteriorated overall. For example, a site with a moderate defect running along its entire length would have a value of 100% for the proportion of the site classed as red or amber. The average score of the site however would be quite low as the defect would not be serious enough to score particularly highly. Both aspects are relevant to the assessment of road condition, therefore both measures were considered and compared with the engineers’ assessments.

To determine which of the parameters had the strongest influence on the engineering assessment results the model was run several times on the same dataset, but with the weightings set such that only one parameter was being used to calculate the Defects Index in any run. Knowing which of the parameters was most closely related with engineering opinion on site condition enabled us to adjust the weightings of the parameters in the final model to best reflect the views of the engineers.

The reference dataset provided by the engineers also enabled us to implement and test some variations to the model as suggested by CBC and the DIWG, such as the different methods proposed for combining the weighted parameter scores in subsection to obtain an overall score for each 10m subsection.

The use of non-linear weighting curves was investigated to determine whether or not there was any benefit in the use of such an approach. This was done by considering the contribution of a single parameter in isolation, in a similar way to that used to determine which parameters were most strongly related to the engineers’ opinions. The parameter was used with a linear weighting process, and a non-linear weighting curve, and the results compared against the engineers’ ratings.

### 3.3 Network level analysis

In addition to the comparisons of the model results with the engineers’ assessments the work also included processing all the available data from a number of Local Authority networks, to determine whether the proportions categorised by the model as Red, Amber and Green were realistic, in terms of the amount of investigation and maintenance that a Local Authority is able to carry out. This was done quite subjectively, but in essence was to check that the model did not produce results which would be unrealistic to act upon, such as having 90% of the network being Green, or 50% being Red. Where possible the network level results were assessed against existing BVPI results.
4 Model realisation

As mentioned in Section 3.1 the Defects Index model was created in Microsoft Excel, using Visual Basic macro commands.

4.1 Input data

The SCANNER data was supplied by the survey contractors, with permission of the relevant Local Authorities, in HMDIF format. This is the same format as the data which the contractors supply to the Local Authorities. These HMDIF files were processed using TRL’s ChartCrack software to produce output files in a format that could be easily imported to the model. The ChartCrack output files contained values every 10m for nearside and offside rut depth; 3m and 10m LPV; whole carriageway cracking; nearside and offside wheeltrack cracking; and texture depth. The model was also designed to accept information regarding the road class (A, B, C or unclassified) and environment (rural or urban) when available.

4.2 Model operation

As implemented, the model performs the following operations:

**Threshold and weight data**

For every piece of input data the measured parameter value is compared against predefined threshold values. These thresholds vary according to road class and environment. If the input value is below the lower threshold ($T_{\text{LOWER}}$) it is assigned a weighted value of zero. If the input value is above the upper threshold ($T_{\text{UPPER}}$) the parameter is assigned a weighted value of 100. If the input value is between $T_{\text{LOWER}}$ and $T_{\text{UPPER}}$ it is assigned a weighted value between 0 and 100, depending on the value of the input parameter, and the form of weighting which is being applied.

Figure 1 illustrates this weighting procedure, and demonstrates the use of linear and non-linear weighting curves.

![Figure 1: Example of linear and non-linear weighting curves as applied to rutting measurements](image-url)
Each parameter score was then adjusted by two weighting factors representing the reliability of the measurement of the parameter, and the importance placed upon the parameter in decisions regarding the condition of a site. For example, rut depth measurements are considered more reliable than cracking measurements and so the reliability factors placed more weight on rutting as a parameter. Short wavelength unevenness is also considered more important to the engineer than medium wavelength unevenness and so the importance factors placed twice as much weight on 3m LPV values than they did on 10m LPV.

**Combine weighted data**

The weighted data was combined in one of a variety of ways in order to produce a single score value for each subsection of the dataset. The combination methods tested were:

- Sum of all weighted parameter scores;
- Average (mean) of all weighted parameter scores;
- Maximum of all weighted parameter scores;
- Order weighted average of all weighted parameter scores.

The order weighted average was calculated by weighting the individual parameter scores, then ranking them in terms of magnitude. The order weighted average value was the sum of half of the largest weighted parameter value plus a quarter of the second largest weighted parameter value plus 15% of the third largest and 10% of the fourth largest weighted parameter values.

The use of ‘families’, or groups of similar parameters, where only the maximum of a set of similar parameters would be used in the calculation, was also investigated. For example, was it appropriate to include both whole carriageway cracking and wheeltrack crack measurements when they were possibly providing information on very similar defects?

**Determine condition classification**

The single score value for each subsection was compared against predefined values to determine whether the condition of the subsection in question was categorised as ‘Green’, ‘Amber’ or ‘Red’.

Following the recommendations made by the DIWG a subsection scoring over 100 was categorised as Red, a subsection scoring 100 or less, but more than zero was Amber, and any site scoring zero was classed as being Green.

**Combine condition classifications**

The condition classifications of all 10m subsections were amalgamated for each network section to determine the proportions of each section or site, or the entire network, which were ‘Green’, ‘Amber’ or ‘Red’. Similarly, the average score of each section, site or network could be found by summing the individual scores obtained for each subsection, and dividing by the number of subsections comprising the larger entity.

### 4.3 Initial model settings

The initial threshold values, weightings, importance and reliability factors used in the model realisation were those suggested by CBC on behalf of the DIWG (CBC, 2004), and provided in Table 2.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>units</th>
<th>Road class</th>
<th>Lower threshold</th>
<th>Score at lower threshold</th>
<th>Upper threshold</th>
<th>Score at upper threshold</th>
<th>Importance to engineer</th>
<th>Reliability of measurement</th>
<th>Reliability x Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rutting</strong></td>
<td>mm</td>
<td>A, B (urban)</td>
<td>10</td>
<td>0</td>
<td>20</td>
<td>100</td>
<td>0.9</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (urban)</td>
<td>12</td>
<td>0</td>
<td>25</td>
<td>100</td>
<td>0.9</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (rural)</td>
<td>15</td>
<td>0</td>
<td>30</td>
<td>100</td>
<td>0.9</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td><strong>Texture</strong></td>
<td>mm</td>
<td>A, B (urban)</td>
<td>0.3</td>
<td>100</td>
<td>0.6</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (urban)</td>
<td>0.3</td>
<td>100</td>
<td>0.6</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (rural)</td>
<td>0.3</td>
<td>100</td>
<td>0.6</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td><strong>3mLPV</strong></td>
<td>mm²</td>
<td>A, B (urban)</td>
<td>4</td>
<td>0</td>
<td>10</td>
<td>100</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (urban)</td>
<td>7</td>
<td>0</td>
<td>17</td>
<td>100</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (rural)</td>
<td>15</td>
<td>0</td>
<td>25</td>
<td>100</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td><strong>10mLPV</strong></td>
<td>mm²</td>
<td>A, B (urban)</td>
<td>21</td>
<td>0</td>
<td>56</td>
<td>100</td>
<td>0.4</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (urban)</td>
<td>45</td>
<td>0</td>
<td>90</td>
<td>100</td>
<td>0.4</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (rural)</td>
<td>45</td>
<td>0</td>
<td>130</td>
<td>100</td>
<td>0.4</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>Whole carriageway cracking</strong></td>
<td>%</td>
<td>A, B (urban)</td>
<td>0.15</td>
<td>0</td>
<td>2</td>
<td>100</td>
<td>0.7</td>
<td>0.5</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (urban)</td>
<td>0.15</td>
<td>0</td>
<td>2</td>
<td>100</td>
<td>0.7</td>
<td>0.5</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (rural)</td>
<td>0.15</td>
<td>0</td>
<td>2</td>
<td>100</td>
<td>0.7</td>
<td>0.5</td>
<td>0.35</td>
</tr>
<tr>
<td><strong>Wheel track cracking</strong></td>
<td>%</td>
<td>A, B (urban)</td>
<td>0.5</td>
<td>0</td>
<td>5</td>
<td>100</td>
<td>0.9</td>
<td>0.5</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (urban)</td>
<td>0.5</td>
<td>0</td>
<td>5</td>
<td>100</td>
<td>0.9</td>
<td>0.5</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (rural)</td>
<td>0.5</td>
<td>0</td>
<td>5</td>
<td>100</td>
<td>0.9</td>
<td>0.5</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Table 2: Initial model settings used as suggested by DIWG (CBC, 2004)
5 Data used

Most of the data used in the research was from A roads. This was due to the fact that at the time of the research only the class A roads had been surveyed in earnest. Consequently, the threshold levels and parameter weightings, importance and reliability factors have not been assessed for other road classes.

5.1 Comparison with engineers’ assessments

A number of Local Authorities responded to the documents sent out (Appendix A and Appendix B) requesting their cooperation. These authorities allowed TRL to have access to their SCANNER survey HMDIF files, and also identified and performed engineering assessments on suitable sites from their networks.

Not all the sites which were nominated could be used in the final assessment. Problems with data, with sites being too long to be realistically categorised with a single number, or with inconsistent section naming conventions meant that some data had to be excluded from the final assessment.

In spite of these problems, the final analysis was performed using nearly 140km of data. This data comprised 68 sites with engineering assessment.

5.1.1 Distribution of site lengths in reference data

Figure 2 shows the distribution of lengths among the sites provided for use as reference data. Most of the sites were less than 2000m in length, and so could be reasonably categorised by a single condition indicator value. This is especially true given that the engineers had specifically been asked to provide sites which were in reasonably consistent condition along their entire length and in both directions. In order to remove sites which were not in consistent condition along their entire length, sites longer than 2500m were excluded from the final analysis.

![Figure 2: Distribution of site lengths supplied with reference engineering assessments](image)

5.1.2 Distribution of site ratings in reference data

Figure 3 shows the distribution of assessment values given to the reference data sites by the engineers. The engineers were asked to rate each site between ‘sound’ (zero), and ‘having a need for urgent maintenance’ (10). It is apparent from Figure 3 that the sites include a good range of condition, as rated by the engineers. Ideally there should have been an even distribution, i.e. with the same number of sites in each category but the data provided has achieved a reasonable spread and is not dominated by particularly good or particularly poor sites, and as such is acceptable.

Clearly, as the engineers’ assessments are based on subjective opinions there are obvious questions regarding the consistency of the ratings. For example would a site rated as a 2 by one engineer be the same as a site rated as a 2 by another engineer? Overall, it was considered that while there will be
subtle differences between how different engineers would have assessed the same sites, it is unlikely that there will be substantial differences, for example of one engineer rating a site as a 0, 1 or 2, while another engineer rated the same site as an 8, 9 or 10.

![Figure 3: Distribution of site ratings (0-10, 0 = new, 10 = urgent maintenance) supplied by engineers](image)

**5.1.3 Occurrence of specific defects, and seriousness of defect in reference data**

The engineers were asked to specify the presence and severity of certain defects on their nominated sites. The defects chosen were rutting, carriageway cracking, wheeltrack cracking, profile problems such as bumps or unevenness, and non-skidding related texture defects such as fretting or fatting. These defects were chosen as these are the types of defect capable of being detected by SCANNER surveys.

Table 3 shows that all five defects were common in the sites the engineers had assessed, and that all defects were considered a problem, or a major problem on at least 15 sites. It can also be seen that rutting or wheeltrack cracking was very rarely a major problem in sites selected by engineers. Where significant defects were present they were usually carriageway cracking, texture and profile defects.

<table>
<thead>
<tr>
<th>Defect present</th>
<th>Present and a problem</th>
<th>Major problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rutting</td>
<td>(no of sites)</td>
<td>(no of sites)</td>
</tr>
<tr>
<td>Texture</td>
<td>40</td>
<td>15</td>
</tr>
<tr>
<td>Profile</td>
<td>51</td>
<td>22</td>
</tr>
<tr>
<td>Wheeltrack</td>
<td>33</td>
<td>15</td>
</tr>
<tr>
<td>Cracking</td>
<td>38</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 3: Frequency of occurrence and severity of defects present on supplied sites in engineers opinion
Table 3 shows that rutting was present on more sites than whole carriageway cracking, yet rutting was only thought to be the main problem on one of these sites, with the whole carriageway cracking the main problem in nine cases.

5.2 Network level assessments

Network level assessments were performed by processing all the available SCANNER data from each authority, and calculating the proportions of the network which were classed as Red, Amber or Green. No specific reference data was required for this stage, although some authorities did supply their existing BVPI figures for comparison.

The data was provided in HMDIF format by the survey contractors, and, as before, was mainly from class A roads, with a few sections of B, C and Unclassified roads.
6 Results and discussion

Following initial testing and consideration, TRL chose to focus efforts on the selection of appropriate weightings and thresholds for the parameters, with a short investigation of the effect of the use of non-linear rutting curves, as proposed by the DIWG. The use of the ‘families’ approach, wherein only the most severe of a group of similar defects was considered within each subsection was therefore not tested in any depth. Additionally, the different methods for combining the weighted parameter scores to determine the overall subsection score were not investigated at all stages once the initial decision to use the sum of all weighted parameter values was made. A summary of the decisions made regarding the options considered is given below:

- The family approach was not pursued except to avoid the double counting of nearside and offside rutting. Therefore, only the maximum rut depth was considered for each 10m subsection.
- Use of the maximum weighted parameter value to determine the condition of a subsection was considered a disadvantage because the road condition as a whole would not be considered. Decisions would therefore be made using less information than was available. This option was therefore rejected.
- The total of all the weighted parameter scores, and the mean of all the weighted parameter scores, are essentially the same metric. In order to reduce the number of calculations required in the model, only the total score of the weighted parameters was considered.
- The use of an order weighted average of the weighted parameter scores offered a valid alternative to the total score, but increases the complexity of the model. Additionally, it was observed in the preliminary testing that the use of this measure reduces the importance, or totally ignores, of less severe defects. This may cause a length of road which was moderately deteriorated in a number of ways to be classed as being better than it was. For this reason, and the increased complexity of the model, this option was discarded at an early stage in the testing.

6.1 Comparison with engineers’ assessments

As described in Section 3.2.3, to determine which of the parameters used in the model was most closely linked with the engineers’ assessments the model was run several times on the same dataset, but with the weightings set such that only one parameter contributed to the calculation of the Defects Index in any run. To achieve this, the weightings of the other parameters were set to zero.

For each parameter, with its own weighting factor set to be one, and the weights of all others set to zero, the model was run on the reference data sites. The percentage of each site within the dataset categorised as Red or Amber, and the average (mean) score obtained per site were compared with the engineers' assessments. It was hoped that this analysis would show which of the parameters most closely mirrored the engineers’ opinions of the condition of a site.

6.1.1 Results using default settings.

For the testing of each parameter, the threshold settings used were those outlined in Table 2. The model used a linear weighting curve for all parameters. The family approach was used for rutting only. A subsection was classed as Amber if it scored more than zero points, and was Red if it scored more than 100 points.

Figure 4 shows the percentage of each site which was classed as Red or Amber when considering each of the model input parameters in isolation, plotted against the site assessment scores provided by the Local authority engineers. Similarly, Figure 5 shows the mean score for each site, based on the individual scores per subsection, while considering each parameter in isolation.
Figure 4: Percentage red or amber in each site when considering one parameter at a time

Figure 5: Average score for each site when considering one parameter at a time
It was found that the two LPV parameters (3m and 10m LPV) appeared to be most closely related to the engineers’ opinions of the site conditions. This was the case both when using the percentage Red or Amber as a score metric, and when looking at the average score of the site. The next most important parameter appeared to be whole carriageway cracking.

6.1.2 Effect of changing weightings

Figure 6 compares the best fit lines for the relationships found between the average score of each reference site and the engineers’ assessments of the site condition for each model input parameter in isolation. Figure 7 shows the relationships between the proportion of each reference site classed as Red or Amber and the engineers’ ratings of the sites.

We can see from Figure 6 and Figure 7 that 3m LPV and 10m LPV appear to most strongly agree with engineering judgement, followed by cracking. Rutting and texture show no strong relationship with the engineers’ assessment ratings.

The Defects Index model settings proposed by the DIWG (Table 2) emphasise rutting most heavily in the calculations (a Reliability of 1 and an Importance of 0.9, giving an overall weighting of 0.9). 3m LPV is weighted with a factor of 0.8, while 10m LPV has a weighting of 0.4. Whole carriageway cracking is weighted with a factor of 0.35 (0.7 for Importance, but 0.5 for Reliability). It appears, following the analysis performed by TRL, that too much emphasis is being placed on rut measurements in the calculation of the Defects Index. Similarly, it appears that 10m LPV and whole
carriageway cracking should be weighted more strongly. The weighting of 0.8 applied to 3m LPV measurements seems about right.

By altering the model to give more emphasis to 10m LPV values, and whole carriageway cracking, and to give less prominence to rut measurements we can see the effect this has on the results.

Figure 8 shows a plot of the average score for each site against the condition of the site as assessed by the engineers. The green data were obtained using modified weightings of 0.7 for rutting, 0.6 for 10m LPV, and 0.5 for whole carriageway cracking. The red points were obtained using the default model settings as suggested by the DIWG and given in Table 2.

We can see in Figure 8 that the modified weightings for rutting, 10m LPV and whole carriageway cracking result in average scores which show a greater increase as the condition of the site deteriorates. The difference between the average score obtained on each site using the different weightings is less obvious on sites considered by the engineers to be in sound condition than on sites in poor condition.

6.1.3 Effect of changing thresholds

The effect of changing the value of the lower threshold limits was investigated for each parameter in isolation. Figure 9 shows how the average scores obtained using only 3m LPV alter with TLOWER set at 75%, 100% and 125% of the original, DIWG proposed threshold levels.

It was found that it was possible to slightly improve the strength of the relationships between the average score of each reference site, and the proportion of each reference site classed as Red or Amber, and the engineers’ opinions by adjusting the threshold values. For example by lowering the threshold limit more 10m subsections are counted as being defective in terms of 3m LPV. It can be seen in Figure 9 that this results in an overall rise in the average score for each site. This rise in score is not uniform and is more apparent on sites which have been scored highly by the engineers.

However:

- The thresholds for texture and rut depth were not investigated in any detail as they have been set for safety reasons.
• In light of the existing engineering reasons for the threshold levels being set as they are, the
  evidence was not considered a strong enough justification for altering any of the thresholds

![Figure 9: Variation in average score for each site set when considering 3mLPV and varying the lower
threshold level as indicated](image)

6.1.4 Effect of non-linear relationship for weighting rut depth

No evidence was seen that there was any benefit in the use of a non-linear weighting curve. Due to the
increase in the model complexity and the lack of any evidence that it was beneficial the model was
altered to use linear weighting curves for all parameters.

6.2 Network level assessments

At a network level, the condition of a given network can be judged by the amount of Red or Amber
subsections it contains. Figure 10 shows the Red, Amber and Green proportions of various Local
Authority networks, following the application of the recommended Defects Index model.

It can be seen that several of the networks tested contain significant amounts of Red or Amber
subsections. These would be the areas of the network which would be highlighted as requiring
investigation and possible maintenance attention. However, the number of such sites would be
impractical to investigate of maintain in many cases.

![Figure 10: Proportions of selected Local Authority networks categorised as Red, Amber or Green using
final recommended Defects Index model settings](image)
By setting the Amber threshold at 20, which is the score achieved with the least significant defect being half as bad as its upper threshold, the proportion of Amber in each network is significantly reduced. This can be seen in Figure 11.

![Proportions of Green, Amber or Red with Amber threshold set at 20](Image)

**Figure 11:** Proportions of selected Local Authority networks categorised as Red, Amber or Green using final recommended Defects Index model settings, and an Amber threshold of 20.

However, where a single defect results in a score of 19, it could be considered that a Green rating provides an over optimistic view of the condition of the network. Therefore, it may be more appropriate to keep a zero threshold to indicate those sites which are genuinely in sound condition and to have a new threshold at 20 to differentiate between those sites with less severe deterioration and those which were more seriously defective.

This could introduce a new 4 class system, with three threshold levels as follows:

<table>
<thead>
<tr>
<th>Subsection score</th>
<th>Colour</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Green</td>
<td>Sound</td>
</tr>
<tr>
<td>$0 &lt; x \leq 20$</td>
<td>Light Amber</td>
<td>Some minor deterioration present, but not a priority for further investigation</td>
</tr>
<tr>
<td>$20 &lt; x \leq 100$</td>
<td>Dark Amber</td>
<td>Some deterioration present and in need of investigation to determine maintenance need</td>
</tr>
<tr>
<td>$x &gt; 100$</td>
<td>Red</td>
<td>Poor overall condition likely to require planned maintenance soon</td>
</tr>
</tbody>
</table>

**Table 4:** Proposal for 4 level categorisation system
7 Recommendations

Based on this research the following changes are recommended to the implementation of the Defects Index model put forward by the DIWG:

- Use linear weighting processes for all parameters
- Only consider the maximum of the NS and OS rut measurements
- Adjust the weightings of 10m LPV and whole carriageway cracking upwards
  - Recommend 0.6 for 10m LPV, and 0.5 for whole carriageway cracking.
- Adjust the weighting given to rutting downwards to be less influential
  - Recommend a value of 0.7 for rutting – However, the DIWG have indicated they wish to keep the weighting of rutting unchanged (0.9)
- These adjustments to be made by altering the Reliability and/or Importance values
- Use the total score of the weighted parameters every 10m subsection to calculate subsection categorisation
- Introduce a new threshold level at 20, below which subsections are said to be Light Amber. These Light Amber sites will not contribute to the overall condition indicator for the network, but are not ‘Green’

The resulting parameter values are shown in Table 5.

<table>
<thead>
<tr>
<th>Defect</th>
<th>units</th>
<th>Road class</th>
<th>Defect Value</th>
<th>Weight</th>
<th>Defect Value</th>
<th>Weight</th>
<th>Importance</th>
<th>Reliability</th>
<th>Reliability x Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rutting</td>
<td>mm</td>
<td>A, B, C, U (urb)</td>
<td>10</td>
<td>0</td>
<td>20</td>
<td>100</td>
<td>0.9</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (rur)</td>
<td>12</td>
<td>0</td>
<td>25</td>
<td>100</td>
<td>0.9</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A, B, C, U (urb)</td>
<td>15</td>
<td>0</td>
<td>30</td>
<td>100</td>
<td>0.9</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>Texture</td>
<td>mm</td>
<td>A, B, C, U (urb)</td>
<td>0.3</td>
<td>100</td>
<td>0.6</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (rur)</td>
<td>0.3</td>
<td>100</td>
<td>0.6</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A, B, C, U (urb)</td>
<td>0.3</td>
<td>100</td>
<td>0.6</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>3mLPV</td>
<td>mm²</td>
<td>A, B, C, U (urb)</td>
<td>4</td>
<td>0</td>
<td>10</td>
<td>100</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (rur)</td>
<td>7</td>
<td>0</td>
<td>17</td>
<td>100</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A, B, C, U (urb)</td>
<td>15</td>
<td>0</td>
<td>25</td>
<td>100</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>10mLPV</td>
<td>mm²</td>
<td>A, B, C, U (urb)</td>
<td>21</td>
<td>0</td>
<td>56</td>
<td>100</td>
<td>0.6</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (rur)</td>
<td>45</td>
<td>0</td>
<td>90</td>
<td>100</td>
<td>0.6</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A, B, C, U (urb)</td>
<td>45</td>
<td>0</td>
<td>130</td>
<td>100</td>
<td>0.6</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>Whole carriageway cracking</td>
<td>%</td>
<td>C, U (urb)</td>
<td>0.15</td>
<td>0</td>
<td>2</td>
<td>100</td>
<td>0.9</td>
<td>0.55</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (rur)</td>
<td>0.15</td>
<td>0</td>
<td>2</td>
<td>100</td>
<td>0.9</td>
<td>0.55</td>
<td>0.5</td>
</tr>
<tr>
<td>Wheel track cracking</td>
<td>%</td>
<td>A, B, C, U (urb)</td>
<td>0.5</td>
<td>0</td>
<td>5</td>
<td>100</td>
<td>0.9</td>
<td>0.44</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C, U (rur)</td>
<td>0.5</td>
<td>0</td>
<td>5</td>
<td>100</td>
<td>0.9</td>
<td>0.44</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A, B, C, U (urb)</td>
<td>0.5</td>
<td>0</td>
<td>5</td>
<td>100</td>
<td>0.9</td>
<td>0.44</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Table 5: Final model settings to be adopted
Furthermore, it will be necessary to:

- Reassess situation next year once new data (edge condition and enhanced transverse profile parameters) are available

- Repeat the work on the selection of appropriate weights and thresholds once data is available from a significant sample of minor roads (classes B, C and unclassified)
Acknowledgements

The work described in this report was carried out in the Technology Development Group of TRL Limited.

The authors would like to thank the many engineers and Local Authorities who provided assistance and data for this project, without whom the work would have been impossible.

References

*TTS Defects Index Preliminary Analysis.* Chris Britton Consultancy, October 2004
Appendix A.

SCANNER Condition Indicator – Determining Parameter Weightings and Thresholds – Local Authority Data requirements

TRL requires the following information from each Local Authority willing to assist in the research to determine the most appropriate parameter weightings and thresholds for calculation of SCANNER Condition Indicators:

- Information regarding 12 sites in each LA network. These may encompass part of a section, or 1 or more complete sections.
- These sites must be between 1km and 5km in length.
- The condition of each site must be similar in both directions, that is, both CL1 and CR1 must suffer from the same defects in similar ways. Sites in which only one side of the road are heavily defective, and the other side is sound are not desirable.
- TTS data must be available for all sites. This will restrict the selection of sites to A roads only at this time.
- We propose the following initial site selection rating system. Please use this when selecting your 12 possible sites:

<table>
<thead>
<tr>
<th>New condition</th>
<th>Local deterioration</th>
<th>Partial deterioration</th>
<th>General deterioration</th>
<th>Urgent maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Please choose 4 sites scoring 0, 1 or 2 – in sound condition,
- Please choose 4 sites scoring 3, 4, 5 or 6 – not sound, but not yet in need of maintenance,
- Please select 4 sites scoring 7, 8, 9 or 10 – felt to be in need of maintenance.
- Sites selected must be in generally similar condition to that which they were in at the time of the TTS survey. This means that recently maintained sites are not desirable.
- For those sites which are felt to be defective, the defects must be theoretically detectable using TTS surveying methodology. Therefore please bear in mind the capabilities of the TTS survey when selecting sites. For example:
  - Suitable defects would include:
    - Wheeltrack cracking
    - Whole carriageway cracking
    - Rutting
    - Roughness / ride quality / longitudinal profile
    - Texture (but not skidding), i.e. fretting, fatting.
  - Unsuitable defects:
    - Edge defects manifesting themselves off the carriageway, such as overriding or stepping.
    - Friction related texture defects such as measured using SCRIM or Griptester surveys
    - Structural defects not present in the surface layers
    - Excessive patching which is neither cracked or sunken.
- For each site selected, please rate the extent of each of the suitable defects noted above using the following scale:
  - 0 – Not present
  - 1 – Present
  - 2 – Present and a problem
  - 3 – Major problem
- Each site must be provided with clear and unambiguous location referencing data to enable TRL, who have no knowledge of your networks, to know where the site is, and in which direction it is referenced. Essential information includes:
Section ID, start node, end node, Length of section, Direction, Lane, Description (from x to y).

- Information must also be provided regarding the road environment (rural / urban, non built-up / built-up).
- To aid with location referencing we will require either OSGR data for each site, giving the start and end of each section, and/or a map showing the start and end points of each site, with a readable scale in OSGR coordinates.

A template sheet is attached showing an example of how TRL would like the data to be provided. (Appendix B)
**Appendix B.**

Template and example data showing level of information required by TRL.

<table>
<thead>
<tr>
<th>Authority</th>
<th>Englandshire</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site</td>
<td>1</td>
</tr>
<tr>
<td>Road name</td>
<td>A37 1/2</td>
</tr>
<tr>
<td>Rough location</td>
<td>3 km west of Great Wibble on the Wold</td>
</tr>
<tr>
<td>section ID(s)</td>
<td>A37 003/1234, A37 003/1235</td>
</tr>
<tr>
<td>Length of site (m)</td>
<td>2497</td>
</tr>
<tr>
<td>Direction</td>
<td>northbound</td>
</tr>
<tr>
<td>Lane</td>
<td>CL1</td>
</tr>
<tr>
<td>Environment</td>
<td>Rural (non built up)</td>
</tr>
</tbody>
</table>

| Start node OSGR - northing | 597451.1 |
| Start node OSGR - easting  | 174581.2 |
| End node OSGR - northing   | 457123.2 |
| End node OSGR - easting    | 156475.3 |

| Map attached? | Yes |
| Site Description | From Junction with Tidwell Street to Junction with Burns Road |
| Site Condition rating 0 - 10 | 7 - General deterioration |

**Defect ratings 0 - 3**

<table>
<thead>
<tr>
<th>Description</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheeltrack cracking</td>
<td>1</td>
</tr>
<tr>
<td>Whole carriageway cracking</td>
<td>2</td>
</tr>
<tr>
<td>Rutting</td>
<td>3</td>
</tr>
<tr>
<td>Roughness</td>
<td>0</td>
</tr>
<tr>
<td>Texture defects</td>
<td>0</td>
</tr>
</tbody>
</table>

**Additional defect information**

| Other defects  | 1 |
| specify        | Uneven ironwork |
| Other surface features | 1 |
| specify        | lots of white lines and road markings |

| Surface type | Surface dressing, High friction surfacing for last 40m of site |