THE FUTURE OF TRANSPORT

# PUBLISHED PROJECT REPORT PPR816

## Development of SCANNER and UKPMS: Task 1 - Consistency of SCANNER data and Task 2 - SCANNER Condition Parameters

E Benbow, A Wright, N Dhillon, M Harrington and K Nesnas

## Report details

| | |
|---|---|
| **Report prepared for:** | UK Roads Board |
| **Project/customer reference:** | TS/TRBO/SER/2012 - A12184244 |
| **Copyright:** | © TRL Limited |
| **Report date:** | March 2017 |
| **Report status/version:** | Draft |
| **Quality approval:** | |

| | | | |
|---|---|---|---|
| Emma Benbow (Project Manager) | | Alex Wright (Technical Reviewer) | |

## Disclaimer



## Contents amendment record

This report has been amended and issued as follows:

| Version | Date | Description | Editor | Technical Reviewer |
|---|---|---|---|---|
| 1.0 | 17/03/2017 | Final draft | EB | AW |
| | | | | |
| | | | | |

| | |
|---|---|
| **Document last saved on:** | 17/03/2017 17:57 |
| **Document last saved by:** | Wright, Alex |

# Table of Contents

# Executive Summary

SCANNER surveys were introduced in 2009 to provide network-wide condition assessment of the local A, B and C road network using survey vehicles that travel at traffic-speed measuring the shape of the road surface using laser sensors, and imaging the surface using digital cameras. The collected data is processed and converted into condition parameters, such as rutting, and delivered in a UKPMS-compliant format to local authorities, for loading into their pavement management systems. It is also used to identify lengths in need of maintenance or further investigation, and to support scheme identification and prioritisation. The data also supports asset valuation, via the Carriageway Condition Index (CCI), which is a methodology recognised by HAMFIG and CIPFA for use in Whole of Government Accounts (WGA) and for reporting within local authorities' own accounts.

SCANNER was developed from the Highways Agency's TRACS survey of the strategic road network. Research supported by the DfT, was carried out between 2003 and 2007 to adopt the survey for local roads. This delivered a range of outcomes, including an updated survey specification, a set of "enhanced" parameters focussed on narrower local roads, and the definition for the SCANNER Road Condition Indicator (RCI), which is used to estimate the overall condition of each length of the network.

In 2014 a development group led by software developers, survey contractors, the SCANNER auditor, and local authorities (the SCANNER Development Group, SDG) commenced a review of the performance and status of the SCANNER survey, in the light of the experience of local authority data users, SCANNER survey contractors and the SCANNER auditor. The groups identified three key areas where enhancements or modifications to SCANNER were required:

- Consistency: Despite the detailed QA and Accreditation process employed for all SCANNER data there continue to be issues identified with the consistency of SCANNER surveys, in particular in the measurement of cracking. (Task 1)
- SCANNER Condition Parameters: SCANNER survey reports a wide range of parameters on surface condition. However, there is concern that these are not well used, and that SCANNER does not report all of the defects that authorities regard as important to include in a condition survey. (Task 2)
- Appropriateness of the SCANNER RCI: Does the SCANNER RCI relate well to LHA maintenance decisions, and how LHAs might want to track the effects of maintenance? Could the SCANNER data be better associated with the treatments that are (or would be) undertaken? (Task 3)

Improvements to data consistency and relevance all improve the value for money obtained from SCANNER surveys. Therefore the Scottish Road Research Board (SRRB), in collaboration with UK Roads Board, commissioned work to investigate and develop SCANNER surveys in the three key areas identified above, which have been separated into Three Tasks. The work described in this report was carried out under Tasks 1 and 2. Task 1 has investigated the consistency of the cracking and rutting data and how it might be improved. Task 2 has investigated if and how the SCANNER parameters can be optimised to reflect LHA needs.

**Task 1:** The cracking data has a significant effect on the year to year consistency of network level reporting. Cracking has been observed to be the main cause of the large inconsistencies seen in the QA audit process. There may be differences between the level consistency of cracking on rural and urban roads, but this was not strongly shown in individual LHAs. However, cracking data collected during the winter months *is* observed to be less consistent than data collected during the summer. Therefore it is recommended that a winter shutdown is implemented, which will require discussion with the survey industry.

There is  currently no method to check that the fleet is consistent in the measurement of cracking, and the repeatability test is also weak. This project has therefore developed enhancements to the cracking Accreditation process. It is recommended that the test for repeatability devised within this project is implemented immediately. A new test for fleet consistency has also been devised within the project. It is a more complex test, that will require experience to understand its effect on the current SCANNER fleet. It is therefore recommended that this test is implemented now and trialled over the next 12 months, to allow SCANNER contractors time to develop an action plan to improve any devices found to be inconsistent. It would become a formal requirement at the end of the trial.

Rutting is generally considered a reasonably consistent parameter. However, whilst inconsistencies tend to be small, they can become significant when combined with other parameters, to influence the RCI. Overall the fleet has become more consistent in the last few years. However, there is a noticeable difference between the fleets of the two current contractors, with an average difference in rut depth of 1.7mm being reported.

Possible routes to improve the consistency of rutting have been investigated that include development in both the collection and the processing technologies. It has been shown that higher resolution systems, with wider measurement width, could provide more accurate and repeatable data. Using a centrally defined and controlled rut algorithm could also improve fleet consistency. As SCANNER contractors now employ such systems (and sample their data down), it should be practical to increase the performance requirements defined in the SCANNER specification. In addition, the TRACS rutting algorithm has been trialled and found that, subject to improvements to both the edge detection algorithm and the placement of the straight edge, it should be able to provide good performance. Therefore it is recommended that  implementation of these updates to the SCANNER requirements should be considered.

**Task 2:** SCANNER delivers more than 20 parameters but only a few are used to calculate the Road Condition Indicator (RCI). Also few LHAs make use of the enhanced parameters provided in the 2007 research. Conversely, the survey does not provide all the condition parameters that are considered to be important by LHAs.  Better value could be obtained from SCANNER if the parameters were optimised to reflect LHA needs. LHAs and PMS providers have been consulted to identify potential revisions/enhancements to the SCANNER condition parameters, or potential new parameters that could be included in a future SCANNER survey. Several observations and recommendations resulted from this consultation and have been to identify a number of potential quick wins (enhancements that could be implemented in the next 12 months) and longer term developments (enhancements that would require a development phase over the next 12-24 months followed by implementation).

- Quick Win 1: Cracking
  - The consistency improvements recommended in Task 1 should be implemented as soon as practical;
  - Of the delivered cracking data, value is being drawn from Whole Carriageway Cracking and Wheeltrack Cracking. The remaining surface deterioration parameters are not required in the delivered data.
- Quick Win 2: Ride Quality
  - Use is only being made of one of the two roughness parameters. LPV should be phased out and replaced with eLPV. This will deliver a more stable and accurate RCI, and will reduce the adverse effect of geometry on the data;
  - The measurement of roughness is failing to report defects present in the offside wheelpath. The measurements from both wheelpaths should be included in the RCI calculation, to provide a more robust assessment of ride quality.
- Longer term development 1: Rutting
  - The improvements to transverse profile recommended in Task 1 should be implemented as soon as practicable. Delivery of wider, higher resolution profile will improve accuracy and repeatability. A replacement for the current rut measure should also be considered. A single rut algorithm across all SCANNER devices would minimise the differences arising from the use of different algorithms by different contractors. The new rutting could be trialled alongside the current rutting, until deemed acceptable;
  - Rut depth is sometimes an inappropriate measure to use on narrow roads (e.g. U roads). Transverse variance would be a more appropriate parameter on these roads. The use of this parameter should be considered further.
- Longer term development 2: Fretting
  - There is a clear call from LHAs for a measure of fretting. The current SCANNER texture variability provides a poor proxy for this.
  - The use of multiple line texture measurements, extracted from high resolution transverse profile data, shows promise for the identification of fretting. A method should be developed to deliver fretting from this data,
- Longer term development 3: Bump/pothole measure
  - There has been a strong request for potholes to be included in SCANNER. The current SCANNER Bump Measure does not provide a reliable network level indicator of the extent to which the network is affected by such features.
  - High resolution transverse profile data could be adopted to provide full lane width longitudinal profile data, from which a more reliable bump/pothole measure could be obtained. Development of this parameter is recommended.
- Longer term development 4: Training
  - There is a need to develop an education strategy. This could be developed alongside the recommendations of Task 3, to include the survey, its measurements and the uses of the data (RCI/UKPMS).
  - The purpose of the strategy will be to develop local authority confidence and expertise in the use of SCANNER data. It is envisaged that delivery is likely to be via high-quality multimedia education materials so that the courses are inclusive and accessible regardless of location, time constraints or other local limitations.

# 1 Introduction

The SCANNER survey provides network wide condition assessment of the local A, B and C road network using traffic-speed survey devices that collect data on the visual condition and shape of the road surface. The collected data is processed and converted into condition parameters, such as rutting and cracking, and delivered in a UKPMS compliant format to local authorities, for loading into their pavement management systems.

The data is used within UKPMS compliant systems for reporting the condition of classified local authority roads. It is also used to identify lengths in need of maintenance or further investigation, and to support scheme identification and prioritisation. The data also supports asset valuation, via the Carriageway Condition Index (CCI), which is a methodology recognised by HAMFIG and CIPFA for use in Whole of Government Accounts (WGA) and for reporting within local authorities' own accounts.

SCANNER (initially called TTS) was developed from the Highways Agency's TRACS survey of the strategic road network. TRACS was designed for condition measurement on roads that were typically wide and even, and with few extremes of geometry. Therefore development was undertaken to adopt the survey for local roads. A programme of research, supported by the DfT, was carried out between 2003 and 2007 to undertake this development. The primary outcomes were revisions to the data collection requirements to better suit local roads, and the delivery of parameters better focussed on narrower local roads, describing defects such as unevenness and edge deterioration. The research also delivered the definition for the SCANNER Road Condition Indicator (RCI), which estimates the overall condition for each length of the network.

SCANNER surveys are governed under the RCMG, and its sub groups. A working group led by SCANNER contractors and the SCANNER auditor (SCANNER contractor liaison group, SCLG) provides a forum for management and review of the on-going accreditation and QA process. A development group led by software developers, survey contractors, the SCANNER auditor, and local authorities (the SCANNER Development Group, SDG) provides a further forum for the identification of any issues that might be present in SCANNER/UKPMS. In 2014 these groups commenced a review of the performance and status of the SCANNER survey, in the light of the experience of local authority data users, SCANNER survey contractors and the SCANNER auditor. The groups identified a number areas where enhancements or modifications to the SCANNER process were required, in particular the following three key areas.

***Optimising the consistency of SCANNER data***

As an important survey for both local and national condition assessment a need for consistency and quality control was recognised from the beginning of the SCANNER process. The SCANNER specification requires that all survey devices are accredited, and includes detailed requirements for external independent auditing of the data delivered to Local Highway Authority (LHA) clients. However, even with this process there continue to be issues identified with the consistency of SCANNER surveys. Of the current core data, cracking is the parameter that raises most concern. It is inconsistent across the fleet of SCANNER devices, in that the absolute intensities of cracking reported differ across the fleet and there is inconsistency in the ability of the devices to report cracking at the same locations. Although rutting is more consistent than cracking, concerns were raised over this measure because issues had been identified with localised bias, noise and inconsistency from device to device (which may be site dependent). This issue is relevant because of the more significant contribution that rutting makes to the SCANNER RCI.

***The SCANNER Condition Parameters***

The SCANNER survey reports a wide range of parameters including texture, ride quality, rutting, cracking, edge deterioration etc. A number of these were introduced at the conclusion of the 2009 research, but there has been no follow-up work to investigate their capability and relevance. There is also concern that SCANNER does not report all of the defects that authorities regard as important to include in a condition survey. For example, surface defects such as fretting, fatting and, perhaps, potholes. The question has therefore been raised as to whether the current parameter set is appropriate or sufficient to support maintenance operations.

***The Appropriateness of the SCANNER RCI***

The review questioned whether the current method of reporting SCANNER data (RCI) matches how Local Highway Authorities (LHAs) make maintenance decisions or how LHAs might want to track the effects of maintenance. Although the RCI reports the percentage of the network that is estimated to be in poor condition (i.e. in a “red” category), this does not necessarily mean that this is the length that needs treatment, or is the length that will actually receive treatment. This reduces the link between the SCANNER data and the LHA maintenance activities. It has been suggested that more value might be obtained from SCANNER if the data could be better associated with the treatments that are (or would be) undertaken.

Thus the Scottish Road Research Board (SRRB), in collaboration with UK Roads Board, have commissioned work to investigate and develop the SCANNER survey. The research consists of 3 tasks, relating to the three key areas identified above:

- Task 1 – Consistency of SCANNER data
- Task 2 – SCANNER Condition Parameters
- Task 3 – Appropriateness of the SCANNER RCI.

This report describes the work carried out within Task 1 and 2, and the recommendations arising from this work. Task 3 is discussed in a separate report (Cartwright & Spong, 2017).

# 2 Task 1 - Consistency of SCANNER Cracking data

Task 1 investigates the consistency of the SCANNER data, focussing on the measurement of rutting and cracking, which were identified as key consistency concerns in the SCANNER Development Group review. This section discusses the measurement of cracking.

## 2.1 Approach

SCANNER is delivered to an "end result" specification, which does not state the method with which cracking should be identified on the local road network. The specification defines the accuracy requirements for the measurement, and how it should be reported in an HMDIF file (percentage of road surface affected). This allows contractors to use any suitable technology and, in theory, allows developments in the field of crack detection to be available for the SCANNER survey. However, this approach derives from TRACS surveys, for which the Highways Agency (now Highways England) would commission a single contract over a long (5 year) period. It has some weaknesses where there are multiple vehicles using different approaches. The flexibility in the performance requirements potentially allows individual devices to achieve accreditation, but with differences occurring between devices in the fleet (in terms of the absolute levels of cracking reported). This has become more of a problem as additional devices have been introduced to the survey.

The method used by current SCANNER survey contractors, to identify cracking on the network, is to collect downward facing images of the pavement surface and then use a computer algorithm to analyse the images to identify the cracks present. The image systems are different between contractors and can differ within individual contractors' fleets. Also each contractor uses their own bespoke algorithm to analyse the images. Potential for inconsistency can arise from differences in the image collection systems used, in addition to differences due to the different analysis methods.

Improvements to cracking consistency could potentially be achieved by developing and then specifying the specific equipment and algorithms to be used for SCANNER surveys. However, this was not considered practical. Such development would be far outside the scope of the project (industry has been working at this problem for over a decade, and yet concerns over consistency still exist), it would also be a fundamental change to the end result approach of SCANNER, and it could result in a mature and significant survey industry (there are 15 current vehicles) having to be significantly updated/replaced.

Therefore, the focus of Task 1 has been to obtain a better understanding of the consistency, in terms of its significance to SCANNER, and has then investigated whether improvements to the Accreditation process could be used to assist in increasing consistency across the current fleet, and any new systems that might join the fleet.

The investigation carried out to support this work has required the collation of large datasets from the network survey and from the accreditation tests. These have then been analysed to understand the issue and to propose solutions. The detail of this work is presented in Appendix A. The following sections present a summary of the results and recommendations.

## 2.2 The effect of cracking consistency on network surveys

The problems seen with cracking consistency have been assessed by using network data. We have considered whether the issue actually has a material effect on network reporting and, if so, how significant it is. Also, whether there is any evidence that any particular aspect affects the performance e.g. road type, urban/rural, survey dates.

### *2.2.1 Effect of cracking on RCI*

The RCI combines the rutting, roughness (LPV), texture and cracking data to obtain an overall score that is used to report the condition of each 10m length. The UKPMS rules and parameters define the thresholds and weightings for calculating the RCI. These rules apply a weighting of only 0.6 to cracking. This means that cracking has less influence than other parameters such as rutting, which is weighted at 1 (McRobbie et al., 2007). This means that cracking itself cannot result in a length being reported as “red”, as this requires the length to have a score >100 and cracking can contribute up to 60 points only.

The SCANNER QA process examines all SCANNER data each year and also calculates an Audit Indicator (AI) reporting the percentage of lengths reported as “red” in each year. This is conceptually similar, but not the same as, Single Data List Items 130-01 and 130-02, the national indicators. The QA also examines changes in the AI. As there is expected to be some stability in the data, Authorities where there are significant changes in the AI are investigated to determine if the change is associated with poor data quality.

The research has investigated the effect of inconsistency in the cracking data on the AI by substituting cracking data from one year’s survey in place of the cracking data reported in the precious year’s survey in the same Authority, to show how the stability of the AI would improve if the cracking data was very stable. For the authorities tested it could be seen that inconsistency in cracking was the main cause for the large inconsistency in AI between the two years, and we can conclude that inconsistency in cracking data can and does have a significant effect on network level reporting.

### *2.2.2 Effect of road environment*

To assess the variability in cracking by road environment, 6 years’ data was collated from the national SCANNER database of SCANNER survey data and the average cracking calculated for each year, broken down by road environment (urban/rural/principal/non-principal). Although there will be subtle differences from year to year, it would be reasonable to assume that the average value would remain similar from year to year. Initial assessment of the data showed that, overall, the variability from year to year is greater for urban roads. In theory, this could arise from the greater influence of more challenging features such as reinstatements, ironwork etc. on the crack detection systems. However, further analysis at the individual authority level did not confirm that the consistency is worse in urban areas. Therefore, whilst the overall network assessment indicates a possible difference between rural and urban, this is not strongly shown in individual LHAs.

### *2.2.3 Seasonal variation*

An assessment was carried out to determine whether the time of year that the cracking was collected has any effect on the consistency. Data was extracted from the SCANNER national database for surveys performed in the summer (May to September) and winter (November to February). The cracking data collected during the winter was shown to be more variable, and tests showed this was particular to cracking (the behaviour was not seen in other parameters, such as rutting). Further separation by both environment (urban/rural) and season showed that the greatest variability can be seen in the winter/urban data, but the inconsistency seen in the winter data is not solely due to the increased variability due to the urban lengths.

Therefore it can be concluded that surveying in the winter appears to have a detrimental effect on the consistency of the data. As a result it was proposed that an approach be adopted to minimise the effect of winter on SCANNER consistency. Several approaches were suggested:

- Implement a winter shutdown for the surveys e.g. between December and February inclusive, similar to that for SCRIM;
- Calculate RCI excluding cracking collected during winter months or include an estimate of cracking instead;
- Mark the data as unreliable, thus enabling Local Authorities to choose whether to include the data in the RCI calculation.

A consultation was therefore carried out with stakeholders on these proposals, asking the following questions:

- What are your thoughts on a winter shutdown (i.e. does it sound like a good idea or would it not really affect the way that you currently use the data)?
- Would this have an effect on when you usually receive your SCANNER data and, if so, how?
- If you would like to see it implemented, what increase in cost (either £/km or %) would you find acceptable to ensure better consistency in the data?
- If surveying continued to happen throughout the year, what are your thoughts on
    - Excluding cracking collected during the winter months from the RCI calculation
    - Excluding cracking collected during the winter months from the RCI calculation but including an estimate of the cracking instead (from previous years' data or average value from the local area)
    - Marking the data as unreliable.

The stakeholders consulted included the following Local Authorities: Bristol, Carmarthenshire, Cornwall, Cumbria, Essex, Leicester, South Lanarkshire, and Worcestershire.

Four of the authorities consulted only had summer surveys, so stated that a winter shutdown would not affect them/improve their data. Six authorities gave an opinion (2 who have summer surveys):

- Most felt that a winter shutdown would be a good idea, if it improved cracking consistency.

- One felt that this was a sticking plaster approach and we should just return to using CVI.
- Most felt that ensuring data was collected in the summer would improve the timelines, in terms of receiving the data and being able to generate their programme of works.
- There was general unease at the suggestion to exclude cracking from the RCI calculation, when the data was collected during the winter. The suggestion to include an estimate of cracking from previous year's data was also met with apprehension.
- All would prefer to see no cost increase but accepted cost increases ranging from 1-2% up to 5%.

### *2.2.4 Thresholds used for cracking in the RCI calculation*

The effect of how the RCI thresholds have been set on the perceived consistency of the cracking data has been investigated by simulating the effects of the consistency using data from the national SCANNER database and observing the change in the RCI (see Section A.1.4 in Appendix A). It was found that the thresholds have a minor contribution to the changes in the RCI, but are unlikely to be the major factor in causing cracking consistency to have a large effect on the RCI.

## 2.3 Improving Consistency via the Accreditation process

The current approach to accrediting SCANNER vehicles is clearly defined in the SCANNER specification (SCANNER Specification, Volume 5). In summary, a vehicle is required to collect cracking data on a set of reference sites for which cracking has been measured using manual assessment methods. The machine provides data which is reported as the total area of cracking in each 50m length. The test and the reference datasets are normalised such that the average level of cracking is 1, and each 50m length is then defined as containing high, medium or low levels of cracking. The machine passes the test if it reports a sufficient percentage of lengths that are also reported by the reference as high, medium or low. Note that the test is spatial, in that the same specifically located lengths must be reported as high, medium or low, not just the overall number. In practice this test has been challenging for contractors to meet, so that there has been a degree of pragmatism included in the assessment process since the commencement of SCANNER surveys, with contractors usually having an ongoing Improvement Action Plan (IAP) to increase their statistical performance. This has led to some improvements in the performance of devices, and has actually reduced the variability (as show in A.2.1), but the consistency of the measurement is still proportionately worse than other measures such as rutting.

The normalisation process is included because, historically, there has been difficulty in providing directly similar absolute values to that reported in the reference. Therefore the focus has been on the reliable reporting of "poor lengths", and not on the absolute level. However, this was a much more appropriate test for TRACS, and has weaknesses once a fleet of different vehicles is in use: A fleet of devices that all passed the accreditation could deliver very different levels of cracking.

The accreditation method is also applied to test the repeatability of the data. In this case the data from two runs from a device are taken and individually normalised, with one run then considered the reference and the second considered the test. It is possible that a device could report twice the level of cracking from one run to another but would still pass this test. This would not be considered repeatable for any other parameter and, whilst this scenario has not been seen in practice, there is a need to develop a more appropriate test of repeatability where the actual values reported are compared.

Therefore there is currently no method to check that the fleet is consistent, and the repeatability test is weak. Such tests should be included in the Accreditation and are discussed in Sections 2.4 (repeatability) and 2.5 (fleet consistency).

It is further noted that the accreditation of devices for cracking is carried out on approximately ten sites, which are spread around Berkshire. Four of these sites are predominantly from the trunk road network (M25, M4EB, M4WB and A329M), and six from part of the SCANNER network. The 6 sites on the SCANNER network comprise over 70km of road and are split by rural/urban as shown in Table 1.

**Table 1: Lengths of sites used for cracking accreditation**

| | Length |
|---|---|
| **All roads included in accreditation sites (excluding trunk roads)** | 72.8 km |
| **of which rural** | 60.87 km (84%) |
| **of which urban** | 11.93km (16%) |

Thus the crack sites used for Accreditation contain short lengths of urban roads when compared to rural roads.  Given the observations made above regarding the influence of urban lengths on the consistency of cracking, it may be sensible to include a higher proportion of urban lengths in the tests.

## 2.4 Development of repeatability tests for cracking

Several methods were identified in this specific field with potential for use in assessing either the repeatability or the fleet consistency of SCANNER devices. These, which included the TRACS approach and the SCRIM approach, were reviewed (Section A.2.2 in Appendix A), and it was concluded that none was appropriate for assessing the repeatability of SCANNER cracking in the SCANNER Accreditation tests.

Therefore other standard statistical tests were assessed for appropriateness to test the repeatability of SCANNER cracking data, and two were identified that showed promise: The Confidence interval (CI) and the Coefficient of variation (CV).

The confidence interval is essentially a measure of how precise the data is i.e. how clustered together it is, with smaller CI values indicating tighter clustering of the data. The coefficient of variation is the ratio between the spread of the data and the mean, which again gives an indication of how tightly clustered the data is.  However, the two parameters are subtly different and highlight slightly different forms of inconsistency. Section A.2.3 of Appendix A describes how these parameters are calculated.

There was a need to determine the length of survey data over which to apply the parameters in the assessment of repeatability. The SCANNER parameters are currently reported over 10m lengths, but during Accreditation, cracking is averaged over 50m lengths, to reduce the effect of location referencing errors on the data. However to assess repeatability we need to consider that cracking is quite a noisy dataset, and a short length assessment is probably not appropriate. Therefore a study was carried out to determine when CI and CV become sufficiently stable for use in assessing repeatability (Section A.2.3.3) and the suitable length over which to assess the parameters has been determined to be 500m

### *2.4.1 Applying the CI and CV to determine repeatability*

The CI and CV approach were developed and tested using QA and accreditation data from the vehicles and sites shown in Table 2.

**Table 2: Data available for accreditation testing development**

| Monthly Primary Sites | | SRR1 and SRR2 (SCANNER re-accreditation test routes) | |
|---|---|---|---|
| **Year** | **Contractor** | **Year** | **Vehicles (Contractor)** |
| **2011** | Jacobs, WDM, YottaDCL | 2013 to 2015 | RAV5 to RAV14 (WDM) |
| **2012** | Fugro, Jacobs, WDM, YottaDCL | | Tempest1, 2 and 3 (Yotta) |
| **2013** | Highway Surveyors, WDM, YottaDCL | | |
| **2014** | Fugro, WDM, Yotta | | |
| **2015** | WDM, Yotta | | |
| **2016** | WDM, Yotta | | |

The average value of CI and CV for data from an individual device, from all Accreditation sites, will inform of the general performance of repeatability for that device. The average CI ($\overline{CI_k}$) and CV ($\overline{CV_k}$) for all devices, accredited during 2015, were calculated using the individual CI And CV values from each 500m length, i.e.

$\overline{CI_k} = \frac{1}{n}\sum_{j=0}^{n-1} CI_j^k$ and $\overline{CV_k} = \frac{1}{n}\sum_{j=0}^{n-1} CV_j^k$. These are plotted in Figure 1 for each device accredited in 2015/16.

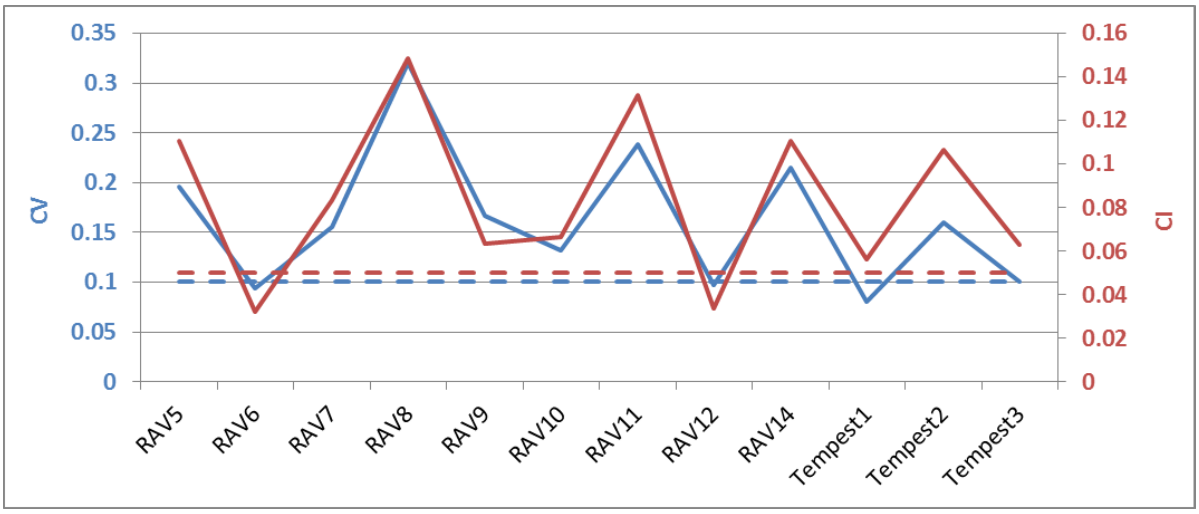



**Figure 1: Average CI and CV for all devices accredited in 2015/16**

The lower a device's $\overline{CI_k}$ and $\overline{CV_k}$ values, the more repeatable the data. In order to determine upper threshold values to apply to these parameters, the 65$^{th}$ percentile of all *CI* values (from all devices, reported over 500m lengths) was calculated. The 65$^{th}$ percentile was chosen since, for normally distributed data, the standard deviation defines a range within which at least 65% of the data lies. The 65$^{th}$ percentile of all *CV* data was also calculated.

Calculating these percentiles for 2015/16 data gives a threshold value of $T_{CIp}$ close to 0.05 for the Confidence Interval and $T_{CV}$ close to 0.1 for the Coefficient of Variation. Therefore these values have been chosen as upper threshold values to apply to $\overline{CI_k}$ and 0.1 for $\overline{CV_k}$, in order to determine general repeatability.

These thresholds are shown on Figure 1 (the blue dashed line for CV and the red dashed line for CI) and, as can be seen, many of the current devices exceed these values, and thus would not pass this test. Hence this will be a tough test to pass, until the repeatability improves. There is scope to reduce the thresholds in future, as the cracking data quality improves.

Table 3 shows the results of applying these thresholds to the $\overline{CI_k}$ and $\overline{CV_k}$ values calculated for the 2015/16 Accreditation data. As can be seen, only two devices pass this test: RAV6 and RAV12. However, it is felt that this single average value test does not really tell you what the device is like in general. The average value can be skewed by outlying values (e.g. large spikes). This may result in the failure of some devices that are usually repeatable but have a spike in the data for one or two lengths. Thus, there is also a need to consider individual 500m lengths for those devices not passing this initial test.

We consider that a device with more than 65% of the lengths having both CI≤0.05 and CV≤0.1 is a repeatable device. Applying this to the 2015 data gives the results in the right hand side of Table 3. As can be seen, all but 4 devices pass this second test.

**Table 3: Average CI and CV values and percentage of lengths not exceeding either threshold (2015 Accreditation data)**

| Vehicle | Average CI value | Average CV value | Pass Test 1? | Percentage of 500m lengths for which CI≤0.05 and CV≤0.1 | Pass Test 2? |
|---|---|---|---|---|---|
| RAV5 | 0.11 | 0.20 | No | 71.6% | Yes |
| RAV6 | 0.03 | 0.09 | Yes | 86.2% | Yes |
| RAV7 | 0.08 | 0.16 | No | 64.2% | No |
| RAV8 | 0.15 | 0.32 | No | 58.7% | No |
| RAV9 | 0.06 | 0.17 | No | 75.2% | Yes |
| RAV10 | 0.07 | 0.13 | No | 71.6% | Yes |
| RAV11 | 0.13 | 0.24 | No | 70.6% | Yes |
| RAV12 | 0.03 | 0.10 | Yes | 84.4% | Yes |
| RAV14 | 0.11 | 0.22 | No | 67.0% | Yes |
| Tempest 1 | 0.06 | 0.08 | No | 82.6% | Yes |
| Tempest 2 | 0.11 | 0.16 | No | 64.2% | No |
| Tempest 3 | 0.06 | 0.10 | No | 58.7% | No |

Inconsistency in cracking data only becomes a problem for the users when large differences in the RCI are seen. For the four devices that do not pass the second test, the inconsistency in their repeat data may not affect the RCI, or there may be specific reasons for localised differences in cracking data that have led to a failure to meet the criteria on a small number of lengths (e.g. due to driving line).

The average contribution made to the RCI by the cracking data for each 500m length is shown for the two test runs in Figure 2 for RAV7. The lengths where the difference in RCI contribution is >10 are arrowed. For this vehicle to be considered consistent, the reason why these differences have arisen will need to be investigated, to determine the cause and the contractor may have to make improvements or focussed requirements may be added to the vehicle's IAP.

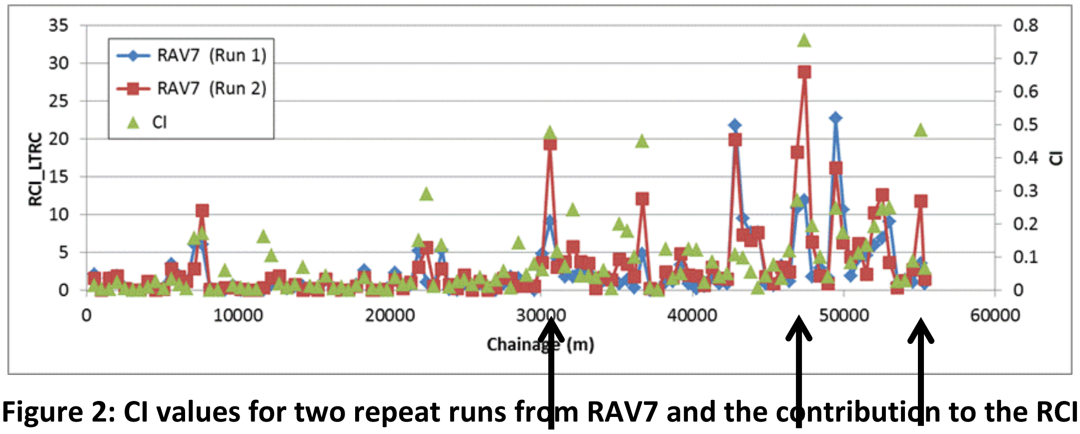



**Figure 2: CI values for two repeat runs from RAV7 and the contribution to the RCI**

## 2.5 Development of fleet consistency tests for cracking

Several methods were identified in this specific field with potential to for use in assessing either the repeatability or the fleet consistency of SCANNER devices. As noted above, these were reviewed (Section A.2.2 in Appendix A) and it was concluded that none was appropriate for assessing the consistency of SCANNER cracking in the SCANNER Accreditation tests

Therefore other standard statistical tests were assessed for appropriateness to test the repeatability of SCANNER cracking data. The data used for this development was the same as that used for repeatability (Table 2).

### *2.5.1 Method to test fleet consistency*

Each device in the SCANNER fleet is required to be re-accredited annually. For most devices this is approximately 12 months after their first Accreditation. Unlike SCRIM testing, in which annual fleet trials are undertaken, the devices do not all get tested on one day – the tests are spread throughout the year. However, although the devices do not perform Accreditation tests at the same time, they do survey the same routes. The auditor monitors these routes regularly, so any significant change in condition are quickly noticed. Thus, for Accreditation it can be assumed that the condition of the routes surveyed will be the broadly same throughout the year and hence the data delivered (rutting, LPV, cracking etc.) should be similar from each device.

The proposed approach to test fleet consistency is therefore based on an assumption that data will be available from the same site for all devices that can be considered broadly comparable. For each device:

- For each 500m length, calculate a representative value for the fleet
- Calculate the bias from the representative value on each 500m length
- Calculate the average bias over the whole site
- If this is less than Threshold X, then the device can be considered to be consistent with the fleet;
- If this is greater than Threshold X, then determine the percentage of lengths where the bias is less than Threshold Y. If this is more than Threshold Z%, then the device can be considered to be consistent.

### *2.5.2 Calculating a representative value for the fleet*

In order to determine a device is consistent with the rest of the fleet, it is first necessary to calculate a value that is representative of the fleet for any reporting length. The simplest calculation would be to take the mean value of the fleet. However, whilst this works well for an evenly distributed fleet (Example 1 in Figure 3), it is less effective when there is one or several outliers in the fleet (Examples 2 and 3 in Figure 3). It also would not highlight when the fleet is split into two groups (Example 4 in Figure 3) and a representative value does not exist for the fleet as a whole.

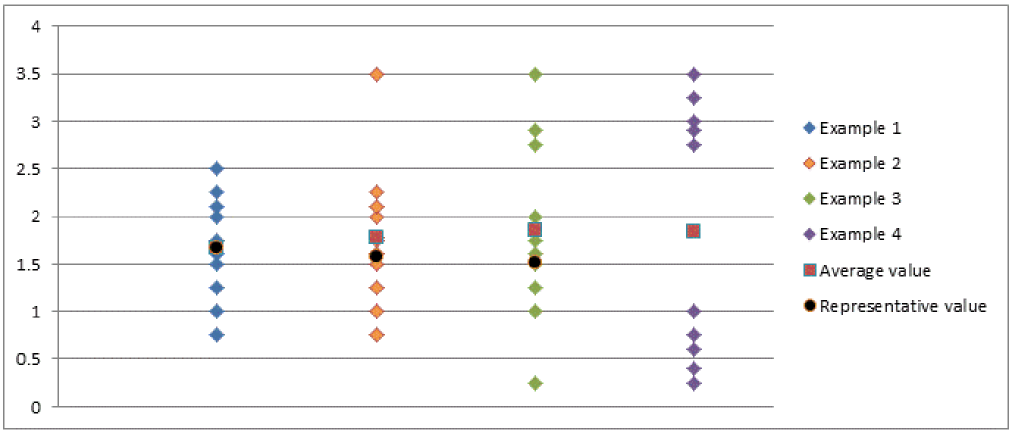



**Figure 3: Examples of fleet distributions (each coloured dot represents an average value for a device)**

It is relatively straightforward for a human to decide which devices should be included in a mean calculation, to determine a value that is more representative of the fleet (black spots in Figure 3). However, this is a subjective calculation which is inefficient, difficult to define in a specification, and open to challenge. Thus there is a desire to calculate a representative value automatically.

Three methods to calculate a representation value were considered: The "centre of gravity" method, the "percentile range" method and the "clustering" method. These are discussed in Section A.2.4.

The "clustering method" was found to be the most robust and consistent of the three methods and thus has been proposed as the most appropriate method to apply for fleet consistency testing. This method determines a representative value by determining which devices report values that are close together (i.e. which devices are clustered) and then calculates a mean of these clustered values.

### *2.5.3 Determining appropriate thresholds for fleet consistency*

There is a need to determine appropriate values for the thresholds X, Y and Z above. As with the CI and CV parameters for repeatability testing, we have calculated the 65$^{th}$ percentile of the absolute biases for all devices, using each of the three approaches for calculating the representative value, described in Section 2.5.2, which gives a value for Y of 0.036 and 65% for Z. As with the CI and CV parameters, it is suggested that the value for X should be the same as Y i.e. X=0.036.

### *2.5.4 Applying the method to Accreditation data*

Rather than having data from all devices collected at the same time, the Accreditation tests are staggered throughout the year. Therefore, the fleet consistency test will have to be applied on a "rolling" basis. The first device to come in for Accreditation in any year would be compared with all other devices holding an Accreditation certificate from the previous 12 months.

To see how this would work in practice, a simulation was carried out using the 2016 accreditation programme. Note this resulted in the exclusion of the Yotta devices from the analysis, since the 2016 Accreditation data was not available for this simulation for the Yotta devices.

The fleet consistency for data from 2015 Accreditations was calculated first and then the 2015 data from individual devices was gradually replaced with 2016 data, to replicate the "rolling" nature of the test. This was used to determine whether the same devices would be reported as consistent with the fleet over time. The results are presented in Table 4. Note that the devices have been added in a random order – not necessarily in the order in which they come in for Accreditation.

As can be seen from Table 4, RAV6 consistently fails the fleet consistency test, whilst RAV7, RAV10, RAV11 and RAV12 consistently pass. RAV14 moves from failing the test to passing, as soon as the new data for this device is considered with the rest of the fleet, suggesting that RAV14 was more consistent with the rest of the fleet in 2016 than in 2015. Similarly, RAV9 fails the test until the 2016 data for this device is considered. Conversely, the 2015 RAV8 data is consistent with the fleet but the 2016 data is not.

**Table 4: Results of fleet consistency tests using clustering method to obtain representative value**

| Vehicle | A: 2015 | | | B: A + 2016 RAV14* | | | C: B + 2016 RAV12** | | | D: C + 2016 RAV9 | | | E: D + 2016 RAV6 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | Pass? | 1 | 2 | Pass? | 1 | 2 | Pass? | 1 | 2 | Pass? | 1 | 2 | Pass? |
| RAV6 | 0.040 | 55.6% | N | 0.040 | 61.2% | N | 0.043 | 57.6% | N | 0.041 | 60.6% | N | 0.044 | 52.0% | N |
| RAV7 | 0.030 | 72.7% | Y | 0.029 | 73.5% | Y | 0.028 | 77.8% | Y | 0.027 | 77.8% | Y | 0.028 | 76.8% | Y |
| RAV8 | 0.031 | 76.8% | Y | 0.031 | 75.5% | Y | 0.029 | 77.8% | Y | 0.029 | 78.8% | Y | 0.029 | 76.8% | Y |
| RAV9 | 0.041 | 62.6% | N | 0.043 | 61.2% | N | 0.040 | 61.6% | N | 0.031 | 71.4% | Y | 0.031 | 73.5% | Y |
| RAV10 | 0.027 | 79.8% | Y | 0.029 | 73.5% | Y | 0.028 | 80.8% | Y | 0.030 | 79.8% | Y | 0.031 | 76.8% | Y |
| RAV11 | 0.035 | 68.7% | Y | 0.033 | 72.4% | Y | 0.037 | 69.7% | Y | 0.034 | 73.7% | Y | 0.035 | 71.7% | Y |
| RAV12 | 0.023 | 78.8% | Y | 0.022 | 82.7% | Y | 0.022 | 81.6% | Y | 0.023 | 81.6% | Y | 0.020 | 85.7% | Y |
| RAV14 | 0.087 | 61.6% | N | 0.060 | 69.4% | Y | 0.058 | 71.7% | Y | 0.058 | 71.7% | Y | 0.057 | 71.7% | Y |

| Vehicle | F: E + 2016 RAV7 | | | G: F + 2016 RAV10 | | | H: G + 2016 RAV11 | | | I: 2016 (+RAV8) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | Pass? | 1 | 2 | Pass? | 1 | 2 | Pass? | 1 | 2 | Pass? |
| RAV6 | 0.044 | 51.0% | N | 0.042 | 54.1% | N | 0.046 | 55.1% | N | 0.043 | 53.1% | N |
| RAV7 | 0.048 | 72.7% | Y | 0.048 | 71.7% | Y | 0.044 | 70.7% | Y | 0.047 | 65.7% | Y |
| RAV8 | 0.029 | 77.8% | Y | 0.031 | 76.8% | Y | 0.029 | 77.8% | Y | 0.098 | 60.6% | N |
| RAV9 | 0.031 | 69.4% | Y | 0.031 | 70.4% | Y | 0.031 | 74.5% | Y | 0.030 | 76.5% | Y |
| RAV10 | 0.032 | 73.7% | Y | 0.034 | 69.7% | Y | 0.035 | 71.7% | Y | 0.034 | 70.7% | Y |
| RAV11 | 0.036 | 70.7% | Y | 0.035 | 71.7% | Y | 0.019 | 87.8% | Y | 0.021 | 85.7% | Y |
| RAV12 | 0.021 | 82.7% | Y | 0.023 | 80.6% | Y | 0.020 | 80.6% | Y | 0.022 | 79.6% | Y |
| RAV14 | 0.055 | 75.8% | Y | 0.055 | 76.8% | Y | 0.052 | 74.7% | Y | 0.055 | 75.8% | Y |

1 = Average Bias for site

2 = Percentage of 500m lengths ≤0.036

* 2015 data for RAV6 to RAV12 and 2016 data for RAV14

** 2015 data for RAV6 to RAV11 and 2016 data for RAV12 and RAV14

#### 2.5.5 Year on year change

Implementing the fleet consistency test on a rolling basis, as each device is submitted for Accreditation (as demonstrated above) should help ensure the stability of average cracking provided by the fleet over time. The method would likely prevent large jumps in the fleet, as was seen between 2013 and 2014 where an overall change of ~0.2% was seen, and subsequently reversed in 2015. This is because any device submitting data that is substantially greater or less than the previous year would fail the test and therefore not be included in any future fleet consistency tests until it is able to provide consistent data.

#### 2.5.6 Potential to provide proxy for reference data

Obtaining reference data via manual analysis of video images of the pavement surface is a very time consuming and expensive task (~1km can be analysed in 1 hour). Thus sites for which reference data is available are usually chosen specifically for cracking analysis and it is not possible to provide reference data for all Accreditation sites, as it is with e.g. rutting.

In calculating a representative value for the fleet, it may be possible to provide a proxy value for the reference, which (unlike reference data from manual analysis) can be directly compared to data from each device (i.e. the data would not need to be normalised before comparison). Thus this approach could be used to provide machine reference data for any site surveyed by all devices in the fleet.

This would enable accuracy testing on a much larger dataset and a much wider range of road types/conditions than is possible at the moment.

## 2.6 Implementation of methods to improve consistency of cracking

The consistency of cracking is an ongoing issue with the SCANNER survey. It is recommended that the following developments identified in this research be implemented:

- Implement a winter shutdown for SCANNER, perhaps from November to January, but this should be discussed and agreed with the survey industry.

- Implement the test for machine repeatability. The repeatability is currently inferred within the Accreditation test and therefore the test developed for repeatability can be considered to be a formalisation of this process. Therefore the test for repeatability should be implemented within Accreditation as soon as possible. This could be achieved by updating the SCANNER specification. An outline of the specification text is provided in Section 2.6.1.

- Implement the test for fleet consistency. This has not been applied before and including it at Accreditation may result in a number of contractor devices failing. It is recommended that the test is introduced but not enforced for at least one year. This will allow a trial period, to identify any issues that need to be ironed out before full implementation, and also to allow the contractors time to determine which of their devices might be inconsistent with the fleet and to develop an action plan to improve this. This could be achieved by updating the SCANNER specification. An outline of the specification text is provided in Section 2.6.2.

### 2.6.1 *Proposed addition to the SCANNER specification to test repeatability in Accreditation*

The following outlines the revisions to the SCANNER specification to incorporate the tests for repeatability into the Accreditation tests:

- Use data collected during surveys of SCANNER Road Routes 1 and 2.
- For each device in the fleet, perform a minimum of 3 runs.
- For each run, compute the average LTRC value over 500m lengths.
- Log transform the LTRC values using log base 10. For LTRC values = 0, set the log transform value to 0.001.
- For each section j for device k
  - Calculate a mean value of all the runs, $\bar{X}_j^k$
  - Calculate the measurement error $\left(\epsilon_j^k\right)^2$
  - Use the data to calculate a confidence interval $CI_j^k$ , defined in section A.2.3.1
  - Use the data to calculate a coefficient of variation $CV_j^k$ as defined in section A.2.3.2.
- Calculate a global $CI_k$ and $CV_k$ are calculated as $CI_k = \frac{1}{n}\sum_{j=0}^{n-1} CI_j^k$ and $CV_k = \frac{1}{n}\sum_{j=0}^{n-1} CV_j^k$.
- A device can be considered to be repeatable if $CI_k$ ≤0.05 and $CV_k$ ≤0.1.
- For all devices that do not meet these criteria, the individual values of $CI_j^k$ and $CV_j^k$ will be assessed. If 65% of the 500m lengths meet the criteria $CI_j^k$ ≤0.05 and $CV_j^k$ ≤0.1, the device will be considered to be repeatable.
- For all devices that do not pass this second stage, we will consider the effect of the repeatability on the RCI:
  - For each run, calculate the contribution of cracking to the RCI for each 10m length: 0 if cracking <0.15%, 100 if cracking >2% and 2000*(cracking – 0.15)/17 otherwise
  - For each run, calculate the average RCI contribution for each 500m length
  - Calculate the difference between these contribution values for each run
  - If the differences are all ≤10, then the inconsistency in the cracking data is unlikely to have an effect on the RCI, as far as the users are concerned and therefore the device would pass the repeatability test.
  - Where differences are >10, investigate the cause of these, to determine whether the device can be considered repeatable.

### 2.6.2 Proposed addition to the SCANNER specification to test fleet consistency in Accreditation

The following outlines the revisions to the SCANNER specification to incorporate the tests for fleet consistency into the Accreditation tests:

- The test would be applied to devices that have passed the repeatability test.
- Use LTRC data collected during surveys of SCANNER Road Routes 1 and 2.
- For each device, compute the average LTRC value for each 500m length, using all data from all runs. This is denoted as $\mu_i^k$ for device k and length i.
- For each 500m length, calculate a representative value for the whole fleet using the average values and the cluster method:
    - Sort the average cracking values, $\mu_i^k$ into ascending order.
    - Starting with the smallest value, look at the difference in value between this and the next larger value.
    - When a difference of >0.1 is found, the values occurring before this gap (and after any previous gaps) are considered to be in the same cluster.
    - Continue to compare adjacent values, until the last value is reached.
    - This will result in between 1 and several clusters being identified.
    - Any cluster with over 50% of the data points lying in it can be considered to be the "representative cluster".
    - If no such cluster exists, inspect the gap size between the clusters. If any clusters are closer than 0.2, then these should be merged to form one cluster.
    - When a representative cluster has been identified, calculate the mean value of all devices in this cluster. This can then be considered to be the representative value for the fleet.
    - If a representative cluster cannot be identified, then the fleet will need to be assessed visually.
- For the test device
    - Calculate the absolute bias of the average for each 500m length from the representative value of that 500m length.
    - Calculate the average bias (i.e. the average of all absolute biases for each 500m length).
    - If the average bias for the whole site is ≤0.036, then the device is considered to be consistent with the fleet.
    - If the bias for the whole site is >0.036, calculate the percentage of lengths for which the absolute bias is ≤0.036. If this exceeds 65% then the device is considered to be consistent with the fleet.

# 3 Task 1: Consistency of SCANNER Rutting data

Although rutting is generally considered a reasonably consistent parameter, the Accreditation and QA process has suggested that there are some inconsistencies in the data, particularly at the lower end of the range (i.e. small values of rutting). Typically the differences fall within the tolerances of the specification (±3mm), and do not affect the current accreditation process. However, these small differences can accumulate with other parameters and influence the RCI. The second part of Task 1 sought to understand these inconsistencies and consider if there would be scope within SCANNER to improve the consistency. These are discussed in this section (3) and in Section 4.

As for the assessment of cracking, the investigation carried out to support this work has required the collation of large datasets from the network survey and from the accreditation tests. These have then been analysed to understand the issue and to propose solutions. The detail of this work is presented in Appendix B. The following sections present a summary of the results and recommendations.

## 3.1 Understanding rutting using the accreditation process

### *3.1.1 Year on year consistency*

The Accreditation data from tests carried out in 2014 and 2015 were examined to determine the consistency with which different SCANNER devices report the same lengths to be in the same RCI category i.e. are lengths reported as Green in 2014 also reported as Green in the 2015 data by a different device. Clearly this assumes little change in the actual rutting, but this is alleviated to some extent by the close attention that the auditor pays to changes in the sites.

It was found that, for some vehicles, significant differences could be identified between the reported categories (B.1.1). This does suggest that the data is inconsistent to an extent that it will affect the RCI calculation. However, overall a very low percentage of the network was found to be affected by lengths contributing to the RCI one year but not the next So, this does not appear to be a large problem on the routes surveyed during Accreditation.

### *3.1.2 Fleet consistency*

A comparison of the average rutting value reported by each device for the last 6 years on the accreditation sites showed that, overall the fleet has become more consistent, and the consistency is good in comparison to cracking (B.1.2). However, there is a noticeable difference between the two current contractors, with Yotta reporting an average rut depth of 1.7mm less than WDM (Figure 4). Since the contractors implement their own rut depth algorithm, this difference could be due to a difference in the measurement of transverse profile between the contractors or a difference in the algorithms implemented.

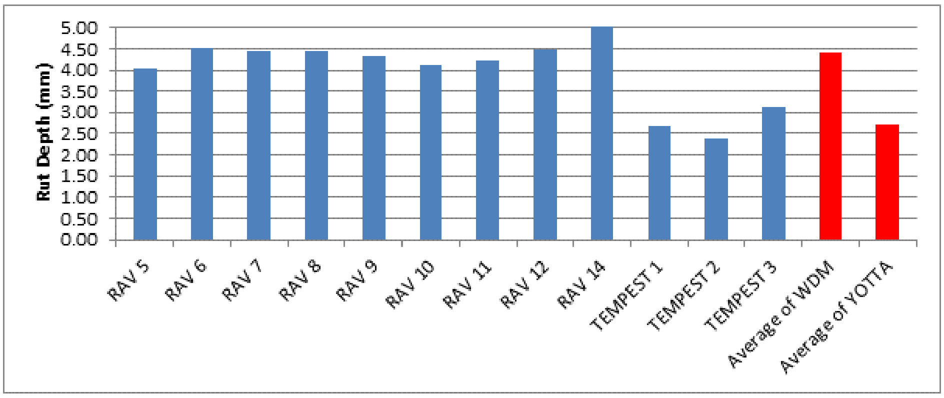



**Figure 4: Average offside rut depths from each device (2014 and 2015 data)**

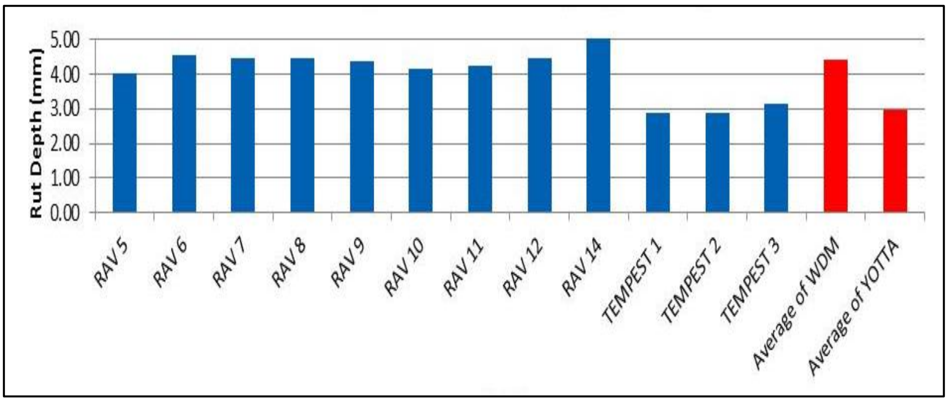



**Figure 5: Average offside rut depths from the fleet, processed using the TRACS rutting algorithm**

Further investigation of the differences found that:

- The differences could be seen even at the site level. The differences affected all levels of rutting, and it is possible that the size of the difference is larger for larger rut depths.
- The differences in rut depth may be caused by different driving lines being taken by the two contractors: the Yotta devices tend to drive further to the left when compared to WDM devices. Thus both contractors are measuring the transverse profile similarly but driving line is causing large differences in these cases.
- The transverse profiles from each device were processed through TRL's bespoke software and analysed using the TRACS rut algorithm to isolate the contractor's rut algorithm as a contributory factor. The difference between the two contractors could still be seen (Figure 5). This suggests that it is not the contractor's algorithms causing the main difference.

The observations suggest that a requirement for a wider profile measurement with higher resolution data in the SCANNER specification could overcome some of these differences.

## 3.2 Understanding rutting using the QA process

The SCANNER QA process examines the distributions of rutting reported in every LHA each year. The process collates the data from the current year and plots it as a distribution and compares it with the data collected in the previous year. The basis of the audit is that, at the network level, an LHA may to expect to have a stable distribution year on year (unless a particularly large maintenance investment has been made). Figure 6 shows the nearside and offside rut depth frequency distributions from Lincolnshire in 2013/14 and 2015/16: The shape of the distributions is not consistent between the two years, and, in this case impacts the value of the RCI calculated (lower threshold for rutting is 10mm – the green dashed lines on graphs).

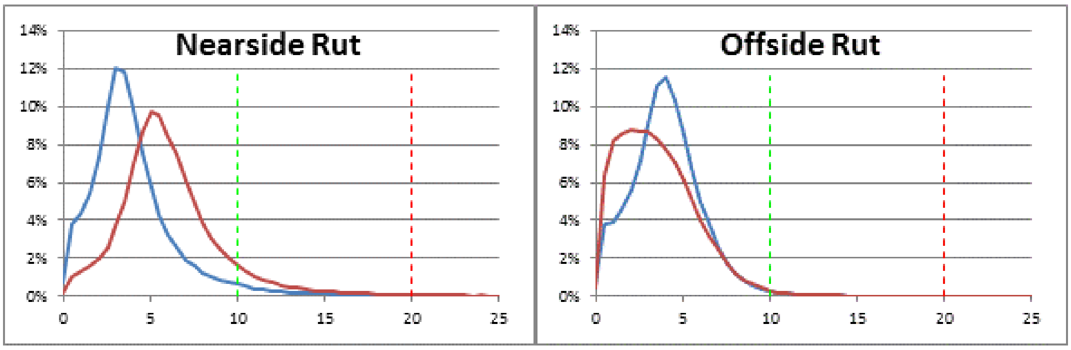



**Figure 6 Nearside (left) and Offside (right) rut depth frequency distributions in Lincolnshire in 2013/14 (red) and 2015/16 (blue)**

There are also many examples, such as those shown in Figure 7, where the distributions are similar in the Amber and Red categories but not in the Green. Whilst the values, in these cases, may not impact the RCI by themselves, the differences can accumulate with other parameters to lead to inconsistency in the RCI.

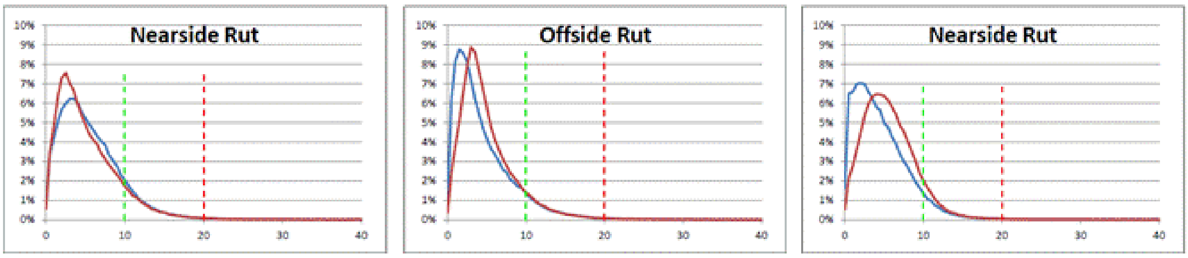



**Figure 7: Examples from QA Audit report of inconsistency of rutting in Green category (rut depths ≤10mm)**

# 4 Task 1: Approaches to Improve Rutting Consistency

The investigation of Section 3 has shown there are some consistency issues with rutting, and that these can affect network level reporting and the RCI. Task 1 has therefore investigated possible routes to improve the consistency via development in collection and processing technologies. A number of investigations were carried out to demonstrate potential improvements, which are explained in greater detail in Appendix B. The following sections present a summary of the results and recommendations.

## 4.1 Use of cleaned rutting

The SCANNER research undertaken in 2007 considered the requirements and challenges that might be presented by the narrower roads found on the local road network. It was particularly noted the presence of edges/embankments on local roads could lead to low quality rut measurements. Therefore work was carried out to develop a new rut algorithm called cleaned rutting. Cleaned rutting is calculated using a centrally defined algorithm that attempts to identify the edge of the road in the data and exclude any points made outside of this edge from the rut depth calculation i.e. the calculation is based on a “cleaned” transverse profile. It was added to SCANNER along with several other enhanced parameters in 2007, but has never replaced the standard rut algorithm.

The use of cleaned rutting to improve year on year consistency was investigated using network data from a LHA. The cleaned rutting (which is provided in the SCANNER dataset) was used to replace the standard rutting within the RCI calculation and in the network audit. Contrary to expectations, a reduction in correlation between the distributions of rutting was performance was seen.

Investigation of this reduction in performance found that this was because the edge detection process within the cleaned rutting algorithm performed poorly. Where the edge was incorrectly detected by the cleaned rutting algorithm there were clear issues with the consistency of the cleaned rutting. Indeed, a brief investigation into the ability of the more recent TRACS rut algorithm to detect the road edge indicated that this had better capability than cleaned rutting.

Thus it seems to be that one of the key things to obtaining an accurate and consistent measure of the rut depths on a road is for the edge to be detected well. This analysis suggests that the current cleaned rutting algorithm would not provide a solution to this problem. Thus it has not been pursued further.

## 4.2 Enhancements in Technology – high resolution profile

SCANNER rutting is obtained by measuring the transverse profile, reporting this as 20 transverse points over 3.2m width, and processing the data through a rut algorithm. This method stems from the technology in place when SCANNER was implemented, where it was likely that the contractor would use 20 individual lasers to measure the profile. Measuring a higher resolution or wider profile was both impractical and expensive. However, there has been a step change in the technology over the last 10 years such that contractors now employ high resolution systems capable of measuring greater than 3.5m width and 100s or

1000s of measurement points. All SCANNER contractors now use this newer technology and sample their data down for SCANNER delivery.

Recognising this change, the TRACS3 contract currently specifies a minimum of 100 transverse points in each transverse profile delivered. TRACS is also required to locate and remove road markings in the rutting calculation. The combination of this and the use of high-resolution systems have improved the rutting consistency in TRACS greatly. Figure 8 shows the differences obtained between two years' rut depths reported in the previous TRACS2 contract (low resolution similar to SCANNER) on 2-way A roads, and the differences obtained between two years' rut depths reported in the current TRACS3 contract (high resolution) on 2-way A roads. There is a much higher percentage of differences <1mm for the high resolution data. This reflects the results seen when moving from 20 point transverse profiles in the TRACS2 contract, to 100 point profiles in the TRACS3 contract.

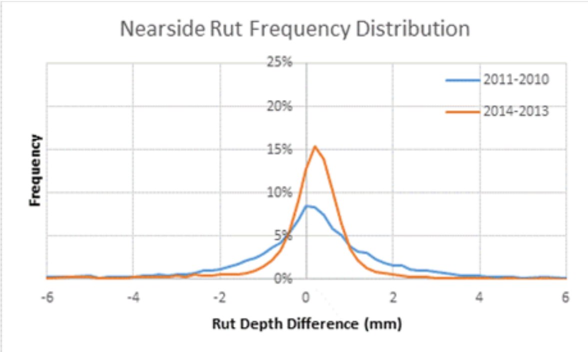



**Figure 8: Repeatability of TRACS data using low resolution (blue line) and high resolution (orange) transverse profile**

The observations above for 2-way Trunk A roads suggest that the introduction of high resolution profile in SCANNER would improve consistency (repeatability) of rut measurement on LA Principal roads, as these are similar in nature.

As TRACS data is not available on local roads, and in particular not on minor roads, we are not able to make such a clear network level demonstration of the potential for high resolution profile on the LHA network data. Therefore, surveys were carried out in this research using HARRIS2, which has a high resolution system, to show the potential for improved consistency on the lower classes of road. A test route was developed which included the SCANNER accreditation sites and an extension to these sites was selected to include challenging narrow roads for which the road edges would be included in the measurement. The surveys measured a 100 point profile over a 4m width and applied the TRACS rutting algorithm. They have shown that:

- Using a high resolution system (and in this case the TRACS rutting algorithm) could provide more accurate data than using a low resolution system – i.e. the measurements are in better agreement with manual reference data (section B.2.2.2).
- Using a high resolution system would also provide much more repeatable data than using a low resolution system.
- There are still challenges remaining in the calculation of the rutting. The higher performance was primarily achieved where the influence of the verge was removed

(manually) from the assessment (section B.2.2.1 and Figure 9). Improvements to the automatic edge detection algorithm would be required, in addition to control over the placement of the straight edge, control over the ability to move the straight edge, etc. (see B.2.3)

- On some narrow roads consideration should be given to using the SCANNER transverse evenness parameter instead of rutting, as this might be more appropriate.

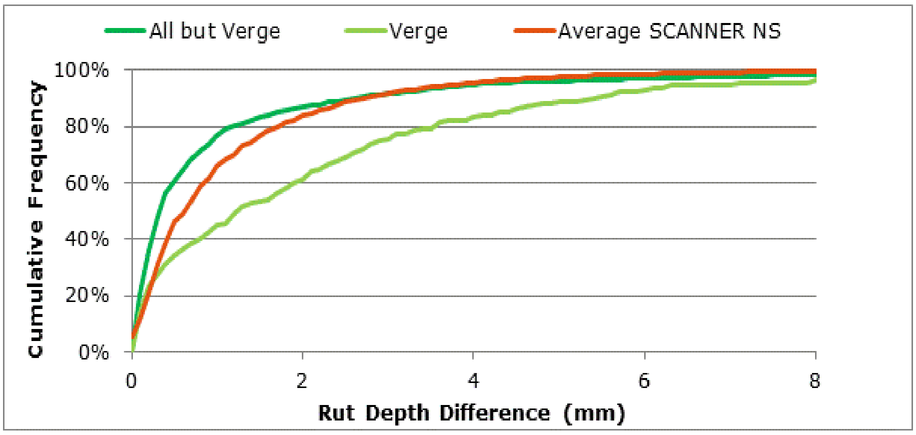



**Figure 9: Cumulative frequency of differences for repeat nearside rut depths: Green lines are rut depths calculated from high resolution data (HARRIS2).**

## 4.3 Implementation of methods to improve consistency of rutting

The consistency of rutting is not such a significant issue as cracking. However, because of its influence on the RCI, there would be benefits in improving the current situation. This research has suggested that rut depths calculated from high resolution transverse profile are more accurate and more repeatable than those calculated from low resolution transverse profile.

- The use of high resolution profile could be implemented via a revision to the specification to require delivery of the enhanced data. It is our opinion that the data should be deliverable by the current SCANNER fleet without replacement of equipment, and therefore should have reasonable cost.
- Time should be allowed to transition to the new data (at least 12 months). This would also allow improvements to be made to the rutting algorithms, in particular the automatic edge detection algorithm and the straight edge placement to optimise the use of the algorithm on lower classes of road.
- The SCANNER transverse evenness parameter might have greater stability on very narrow roads (e.g. U roads) and could be more appropriate for use by those LHAs that commission surveys of these roads. This could be implemented as part of a “U road SCANNER specification”.

# 5 Task 2: SCANNER Condition Parameters - Consultation

## 5.1 Introduction

SCANNER delivers more than 20 parameters but only a few are used to calculate the Road Condition Indicator (RCI). It is also thought that few LHAs make use of the enhanced parameters provided in the 2007 research. Conversely, the survey does not provide some condition parameters that are considered to be important.

Better value could be obtained from SCANNER if we can optimise the parameters to reflect LHA needs. This task has focussed on identifying potential revisions/enhancements to the SCANNER condition parameters, or potential new parameters that could be included in a future SCANNER survey.

LHAs and PMS providers have been consulted to better understand the current use of SCANNER parameters, and the results have been used to categorise the SCANNER parameters into 'valuable/essential'; 'moderate use'; 'worth developing/adding'; 'little use/unreliable'; 'important but not provided' etc. These have then been compared with technical understanding of SCANNER technology, and a review of new technologies, to link the consultation outcomes with potential improvement categories such as: 'quick-win achievable'; 'deliverable (further-research)'; 'remove' etc.

The results of the consultation, and the quick wins identified, are presented in this section. These quick wins have been reviewed with the SCANNER Development Group and further work on them carried out where within the scope of this project. This work is presented in Section 6. Any longer term developments required to achieve the quick wins are discussed in Section 7.

## 5.2 Consultation

The questionnaire sent to stakeholders aimed to determine whether and how the current SCANNER parameters are being used, and what level of importance would be given to each parameter. The stakeholders were also asked for their thoughts on the enhanced parameters (those introduced in 2007 e.g. eLPV, cleaned rutting) and whether they had any additional needs for SCANNER parameters. The questionnaire is given in Appendix D.

The questionnaire was sent to 35 recipients: 29 from England, including 8 Metropolitan/ London Borough authorities, 3 Scottish authorities, 2 Welsh authorities and 1 Northern Ireland authority. In total 15 responses were received:

- 11 responses from English authorities:
  - 7 from counties
  - 2 from unitary authorities
  - 1 from a PFI
  - 1 from London Borough.
- 3 responses from Scottish authorities;
- 1 response from a Welsh authority;
- 1 response from a Northern Irish authority.

## 5.3 Results of consultation

In the questionnaire, the respondents were asked to describe how they used the parameters, what they thought of them and to assign a ranking for importance, with 1 being very important and 5 being not important at all. The results of this are given in Table 5, with the parameters ordered by number of users and then by rating. Some, but not all, respondents gave a rating for parameters that they didn't use, so the average rating shown in this table includes their opinions too.

**Table 5: Number of users of SCANNER parameters and importance given to them**

| Parameter | # of users | Average rating | Parameter | # of users | Average rating |
|---|---|---|---|---|---|
| Rut Depths (nearside, offside) | 12 | 1 | Other Visible Defect | 2 | 4 |
| Cracking (whole carriageway) | 12 | 2 | Transverse/reflection cracking | 2 | 4 |
| 3m LPV (nearside, offside) | 12 | 2 | Transverse variance | 2 | 4 |
| Texture (SMTD) | 11 | 2 | Enhanced 3m LPV (nearside, offside) | 2 | 4 |
| 10m LPV (nearside, offside) | 10 | 4 | Enhanced 10m LPV (nearside, offside) | 2 | 5 |
| Geometry (gradient, crossfall, curvature) | 7 | 2 | Texture Variability (RMST 5$^{th}$ Percentile, 95$^{th}$ Percentile, Variance) | 2 | 5 |
| Edge roughness | 5 | 3 | Edge coverage | 1 | 3 |
| Edge of carriageway cracking | 3 | 3 | Cleaned Rut Depths (nearside, offside) | 1 | 3 |
| Texture (MPD) | 3 | 4 | Transverse unevenness (ADFD) | 1 | 4 |
| Wheel Track Cracking (nearside, offside) | 2 | 3 | Bump Measure (nearside, offside) | 1 | 5 |
| Surface Deterioration | 2 | 3 | RMST Texture depth in the nearside, centre and offside | 0 | 5 |
| Edge steps (at two levels) | 2 | 3 | RMST Variance (nearside, centre and offside) | 0 | 5 |

Looking at the top five, most used, parameters it is not surprising to see that these are the RCI, and original TTS, parameters. Rutting was rated as the most important, with 10m LPV the least important. Also, looking at the bottom end of the table, where there are few users and not much importance is attached to the parameters, it is clear that the Local Authorities are generally not using or are not interested in parameters that weren't included in the original TTS surveys. There are likely to be mixed reasons for this, so each area of measurement has been considered separately in the following subsections.

### *5.3.1 Ride quality: LPV and eLPV*

Almost all of the respondents use the longitudinal profile variance (LPV) parameters. They consider 3m LPV to be quite important, with 10m LPV less so. However, hardly any use the enhanced parameters and have therefore given them a low importance.

It is felt that LHAs probably don't use the eLPV parameters because they already use LPV and the enhanced parameters effectively double up on the original LPV parameters. Thus it would seem appropriate that one of these parameters sets is dropped because they provide fundamentally similar information.

However, although eLPV is the lesser used parameter, we would recommend keeping eLPV, as it has been shown to be a more robust and consistent measure for TRACS. Using it would also align SCANNER with TRACS. Section 6.2 discusses this further.

### 5.3.2 Rutting: Rut depths and cleaned rut depths

A similar argument to LPV could be made in terms of standard and cleaned rutting – users don't use the cleaned rut parameters because they're not in the RCI and they double up the standard rutting parameters. However, unlike with eLPV, the work in Task 1 has shown that cleaned rutting does not perform better than standard rutting.

Therefore there is no benefit in keeping cleaned rutting. In the long term, if a new enhanced rutting measure was provided, this could replace both the rutting and the cleaned rutting parameters. It would be appropriate to report both measures for a number of years, until the new measure was deemed acceptable (running in parallel).

### 5.3.3 Cracking (whole carriageway cracking, edge of carriageway cracking, wheeltrack cracking, surface deterioration) and Other Visible Defects

The whole carriageway cracking parameter, as used in the RCI calculation, is well used and ranked highly. However, the other cracking parameters i.e. wheeltrack cracking, edge of carriageway crack, surface deterioration and transverse cracking, are not used, despite being given a medium importance level. In fact, when the users were followed-up to find out how they were actually using this data, it became apparent that no use was actually being made; the users just thought that they might be useful parameters.

The edge of carriageway crack, surface deterioration and transverse cracking parameters were introduced because of the inconsistency in the cracking measure. If it is possible to make the whole carriageway cracking parameter more consistent (using the developments in Task 1), then these three parameters could potentially be dropped without significant effect.

However, although wheeltrack is not used in any other calculations e.g. CCI, it is used in the treatment rules implemented by UKPMS and would cause a problem if dropped.

Two people reported using the "Other Visible Defect" parameter but when followed-up, it was apparent that this was not the case.

### 5.3.4 Texture: SMTD, MPD, texture variability, RMST texture depth, RMST variance

SMTD is well used but a number of respondents asked why MPD was provided as well, with only 3 people using this parameter.

MPD was introduced because it's a European measure of texture and it was thought that there might be a standard requirement for government to report this measure at some point in the future. This risk may still apply and it would be prudent to continue reporting MPD.

The other texture measures (texture variability, RMST texture depth, RMST variance), which are calculated from measurements made in three lines across the width of the road surface, are not well used. Feedback was also received which said that users had tried to use RMST to get an idea of fretting present but this did not prove very helpful. These parameters could probably be dropped with little effect, but it is noted that no real effort has yet been made to apply them as originally intended.

### *5.3.5 Edge and Bump*

Edge parameters are barely used: Some people stated that they didn't use them because their network is mainly urban, so most of their roads have kerbs etc. However, one urban authority said that they used it where they had cycle lanes, hence the medium level importance rating given. There has also been some investigation in Scotland into the application of these within an edge indicator. It would be recommended that these are kept in SCANNER.

The lack of use of the Bump Measure was quite surprising, since this was developed to identify bump causing features, such as potholes. As discussed in Section 5.3.6 below there was a strong request for potholes to be included in SCANNER. So, there is a need to investigate the behaviour of the Bump parameter to see if it would be useful and stable enough to provide a quick win for potholes/user concerns (Section 6.3).

### *5.3.6 Missing parameters*

The questionnaire also asked the stakeholders what they thought was missing from SCANNER, generating the following suggestions:

- One suggestion for deflection measurements
- One request to have a measure of the change in condition
- Several requests for measures of fretting, potholes, and failed patching.

**Deflection:** The only commercially available equipment that could practically be used to provide network level measurements is the TSD (Traffic Speed Deflectometer). Eight such devices exist in the world currently, with only one of these devices being used in the UK. This TSD is owned by Highways England and used to survey the trunk road network. The measuring equipment of the TSD is placed in the trailer of an HGV and thus would probably be unsuitable for surveys of non-principal roads. However, research has been undertaken to investigate the use of this technology on local roads in the UK, which suggested it could be usable on principal roads (D Wright et al., 2014). However, it would be unlikely to be practical to add this to SCANNER. The most appropriate solution would be to make systems available for commercial surveys on principal roads, perhaps via a principal road network TSD specification, similar to that used by Highways England on strategic roads.

**Change in Condition:** The change in condition calculated is very affected by data alignment and just looking at e.g. the change in rutting on a 10m length can give very misleading results. Research for Highways England, has shown that access to the raw measurement data can be used to align the data suitably so that alignment errors can be overcome and change estimated (McRobbie et al., 2017). Currently the main output of a SCANNER survey is an HMDIF file, which contains only processed data e.g. LPV, rutting values. However, the survey does already have a requirement for the ability to deliver the raw data. This would result in large amounts of data being delivered, but should be manageable using modern IT systems. However, it would require updates to the specification to deliver this data, a process to utilise the measurements and a process to manage it within asset management systems. Due to the complexity of this we have not investigated this further in this work.

**Fretting:** The three lines of texture measurements used so far in SCANNER do not appear to be good enough to give an estimate of fretting, so a quick win for this parameter is not achievable with the current measurements. However, a measure of fretting does exist in the

TRACS which is obtained by processing texture data measured in a minimum of 38 lines covering a width of 3.8m. It is possible that comparable texture measurements could be obtained from high resolution transverse profile systems (i.e. equipment already being used by the SCANNER contractors). The fretting algorithm implemented for TRACS may be appropriate to apply to data collected on the local road. Therefore an investigation of the feasibility of this has been carried out (Section 7.2).

**Potholes and failed patching:** Due to the frequency of SCANNER surveys, the survey would only be able to provide a snapshot of such features on the network. The Bump Measure was developed to identify lengths that contain features that would cause a bump i.e. discomfort to the users. This should be able to identify potholes and failed patch edges occurring in the wheelpaths. However, this measure has been shown to be inconsistent on a length by length basis, possibly because it is very sensitive to driving line. The suitability of the bump measure to identify potholes/failed patching is investigated further in Section 6.3.

## 5.4 Recommendations resulting from the consultation

The following observations and recommendations can be made, following the consultation.

For the standard parameters:

- The most important parameters seemed to be the traditional or "standard" parameters used in the RCI calculation: Rutting, Cracking, 3m LPV, Texture, 10m LPV.
- Only Whole Carriageway Cracking is used, thus other parameters (edge of carriageway crack, surface deterioration, transverse cracking, and other visible defects) could be dropped. Wheeltrack cracking is needed for the UKPMS treatment rules.
- Standard and Cleaned Rutting provide the same data, so one of these could be dropped. As Task 1 has shown that Cleaned Rutting can be unreliable it is recommended that Cleaned Rutting is dropped. Standard Rutting should be enhanced alongside the introduction of high resolution profile.
- The texture parameter, SMTD, is well used and, whilst MPD is not so well used, it is recommended that this should be kept since it is a standard European measure of texture.
- The geometry parameters are well used, due to their inclusion in calculation of site category for assessment of skid resistance. No changes are needed for these.

For the enhanced/new parameters and items considered missing from SCANNER:

- Many users didn't seem to know about or understand the enhanced parameters, which has probably resulted in very little use being made of these (despite them being implemented for nearly 10 years now). Therefore there appears to be a need for education, to ensure that best use of the data available is made.
- eLPV and LPV essentially provide the same data, so one of these should be dropped. We recommend that LPV is phased out and replaced with eLPV.
- The edge parameters are not widely used but have potential for use in an edge indicator. Therefore we recommend that these parameters are kept.

- There is evidence of a need for a measure of fretting. The enhanced texture parameters (RMST, RMST variance, texture variability) that were developed as an initial attempt to identify this defects are not well used and do not meet the users' requirements. Their continued used should be reviewed in the light of undertaking further developments in the measurement of fretting.
- There is evidence of a need for a measure of potholes. However, the existing Bump Measure may not be a strong tool for this. Dropping the measure should be considered in the light of developing a more powerful replacement.

As a result of the above observations we have identified a number of potential quick wins and longer term developments, which are discussed in Section 6 and Section 7 respectively.

# 6 Task 2: SCANNER Condition Parameters - Quick Wins

We define *Quick Wins* as enhancements that could be implemented in the next 12 months. We have identified the following potential quick wins from the results of the consistency work in Task 1, and the changes and improvements identified through the consultation.

## 6.1 Quick Win 1: Cracking

The consistency improvements recommended in Task 1 should be implemented via an update to the SCANNER specification, as discussed in Section 2.6.

Only Whole Carriageway Cracking and Wheeltrack Cracking are used out of all surface deterioration parameters. There seems no benefit in continuing to provide the other parameters in SCANNER. These could be removed from the delivered data.

## 6.2 Quick Win 2: Ride Quality

Experience with the eLPV measure in TRACS has shown that it is a more robust and consistent measure than LPV. There are two areas where this could bring improvements:

- eLPV is expected to provide a more consistent measure of ride quality on roads with varying geometry
- SCANNER provides a measure of eLPV in both wheelpaths, which should provide a more robust assessment of ride quality than the current single wheelpath measurement.

The following subsections investigate whether eLPV is a more robust and consistent measure than LPV and also investigate the extent and size of any change that might be expected for the RCI as a result of using eLPV in place of LPV.

### *6.2.1 Use of eLPV to reduce influence of geometry*

Both LPV and eLPV are obtained by applying a filter to longitudinal profile data to remove long wavelength features, and then calculating the sum of the squares of the filtered profile. They essentially provide the same information. eLPV was introduced to replace LPV in TRACS survey because it had been noted that LPV (particularly the 10m and 30m LPV parameters), is affected by road geometry. Large values of LPV would be obtained on otherwise smooth roads with e.g. high levels of gradient. Having introduced the eLPV measure, it has since been shown to be a more robust and consistent measure for TRACS surveys. If this is also the case for local roads it would be beneficial to keep eLPV and phase out the use of LPV. This would also have the added benefit of aligning SCANNER with TRACS. However, changing to eLPV in the RCI calculation might lead to a step change in the RCI.

The extent of the effect of geometry on LPV on the local road network has been investigated by examining (e)LPV data on the local road network in Devon and on the SCANNER accreditation road routes. The results are presented in detail in Section C.1.1 and summarised here.

On the Devon network there is clear evidence that significantly higher proportions of the road network are reported to be in poorer condition when using LPV. To confirm that this can be linked to geometry the Devon network was broken into lengths classified by geometry and it was shown that the lengths with higher variation in geometry (gradient and crossfall) are reported as rougher by LPV than by eLPV, suggesting that LPV incorrectly associates geometry with roughness.

The consistency of the two measures was also investigated using the SCANNER road routes and QA audit data, where it has been found that

- eLPV is as consistent as, or is more consistent than, LPV on the SCANNER road routes (Section C.1.2).
- eLPV appears to be more consistent when assessed during QA Auditing (Section C.1.3).

### *6.2.2 Step change caused by using eLPV in RCI calculation*

It is likely that switching from LPV to eLPV in the RCI calculation will cause a step change in the national condition indicators, due to the difference in behaviour of the two parameters, especially for lengths where high levels of road geometry are present. To investigate what this step change might be, the change in Audit Indicator (used in the QA Audit reports) has been calculated for several authorities, including:

- Shetlands, Herefordshire and Devon (very rural authorities);
- Bracknell and Blackburn (semi-rural authorities);
- Trafford (metropolitan authority);
- London boroughs;
- Birmingham and Hounslow (urban authorities).

Figure 10 shows the change in the Audit Indicator seen when replacing LPV with eLPV and also the percentage of the network that is urban. As can be seen, using eLPV instead of LPV always results in a reduction in the Audit indicator, ranging from a very small reduction of 0.2% to a large reduction of 4.8%. In general, the change in the Audit indicator is larger the fewer urban lengths contained in the network (i.e. the more rural a network is). This might be expected since higher levels of curvature are often seen on rural roads, compared with urban roads.

These results suggest that an average reduction of about 1.5% would be seen for most authorities.

This is significant and may cause issues if eLPV just replaces LPV. Thus it is recommended that two indicators are provided for the LHAs for the next few years: One calculated using LPV, the other calculated using eLPV. This would enable any step change to be quantified and accounted for.

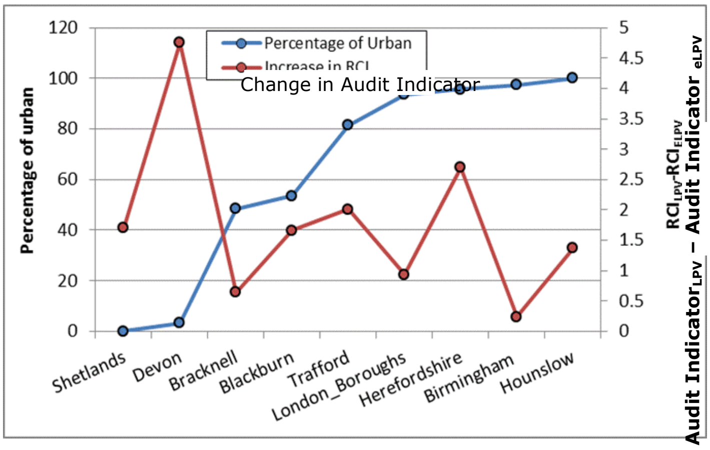



**Figure 10: Decrease in Audit Indictor seen when changing from LPV to eLPV**

### *6.2.3 Including offside eLPV in the RCI*

Currently ride quality is only assessed in the RCI in the nearside wheelpath. However, SCANNER reports longitudinal profile in both wheelpaths. Previous user perception trials of ride quality have suggested that, if only one wheelpath is used to report condition, this will significantly under-report the actual number lengths that have poor ride quality.

It is recommended that the offside LPV/eLPV be included in the RCI, with the calculation using the poorer of the two values (similar to the way rutting data is used). As with the introduction of eLPV into the RCI calculation, introducing offside data is also likely to result in a step change to the Audit Indicator, and the network indicators calculated by the LHAs. This has been investigated for the same LHAs as considered for Section 6.2.2. The Audit Indicators calculated for these using nearside LPV data, nearside eLPV and both nearside and offside eLPV data is shown in Figure 11. The step change seen between the Audit Indicator, calculated using LPV and then nearside and offside eLPV is shown in Figure 12.

Interestingly when using both nearside and offside eLPV a similar AI is obtained to when only nearside LPV is used in the current RCI (Figure 12). This is because using nearside eLPV in the RCI calculation in place of LPV reduces the AI. However, adding in offside eLPV adds additional lengths reported classified as poor and thus increases the AI slightly. However, the introduction of both wheelpaths still results in a change and there may still be significant differences in the values calculated (e.g. Devon, Herefordshire). Therefore it would still be helpful to the users to have a phased in approach of this, as suggested in Section 6.2.2.

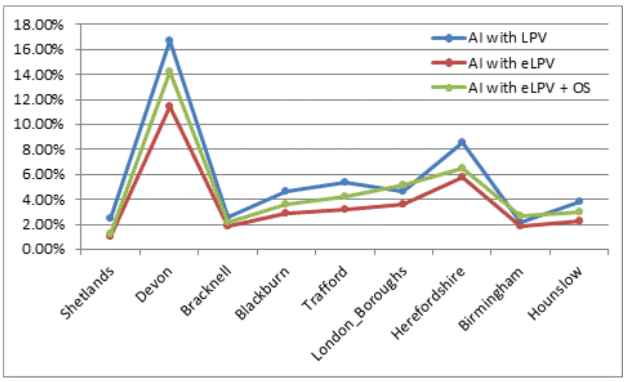



**Figure 11: Audit indicator values for several LHS, calculated using nearside LPV, nearside eLPV and both nearside and offside eLPV**

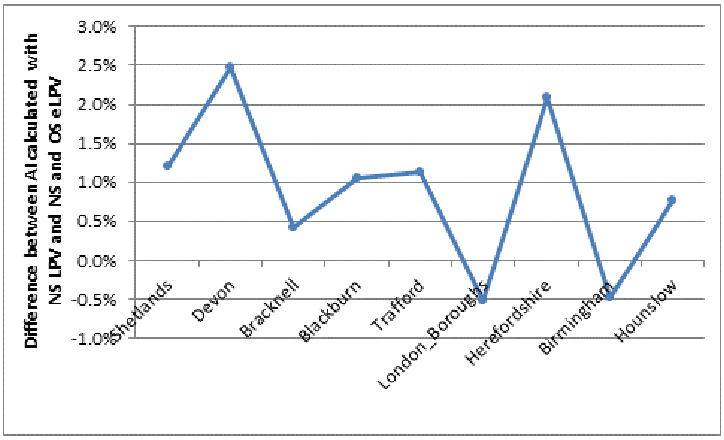



**Figure 12: Difference between the Audit Indicator calculated using NS LPV and NS and OS eLPV**

## 6.3 Quick Win 3 - Bump Measure

*There has been a strong request for potholes to be included in SCANNER. There is a need to investigate the behaviour of Bump and see if it would be useful and stable enough to provide a quick win for potholes/user concerns.* This is investigated further in this section.

The Bump Measure was developed for SCANNER as a result of user perception studies that suggested that, whilst LPV correlated well with the users' opinion of general ride quality, it did not correlate well with discrete ride quality features, which would cause short-lived discomfort i.e. bumps caused by potholes, poorly aligned concrete slabs, failing bridge joints etc. The measure was introduced to SCANNER in 2007.

Feedback from use of the measure in TRACS has suggested that the measure is inconsistent, when considered on a length by length basis. It is thought that this is because the Bump Measure is derived from longitudinal profile, which is only measured in two lines – one in the nearside wheelpath, the other in the offside wheelpath, and thus can be significantly affected by driving line. However, whilst it might not be possible to use the measure on a length by length basis (i.e. to consistently locate bump causing features), it was suggested that it may be able to provide a network level indication of how much of a network is affected by bumps.

To investigate this, the year-on-year reporting of the percentage of lengths containing a bump has been calculated for several local authorities. The authorities considered include examples of mainly rural authorities, mainly urban authorities, mixed authorities, a London borough and a metropolitan authority. The results are shown in Figure 13. As can be seen for some authorities (Birmingham, Shetland), the measure reports roughly the same amount (percentage) of bumps each year. However, for most it is inconsistent with some experiencing very large changes e.g. Hounslow, Bracknell, Trafford. It can also be seen from Figure 13 that, in general, the percentage of lengths containing a bump is less in the offside than the nearside. Reassuringly, splitting the data by road class does report that A roads generally have the least number of bumps, whilst the C roads have most. However, it does not appear that the measure is any more consistent for any individual class of road.

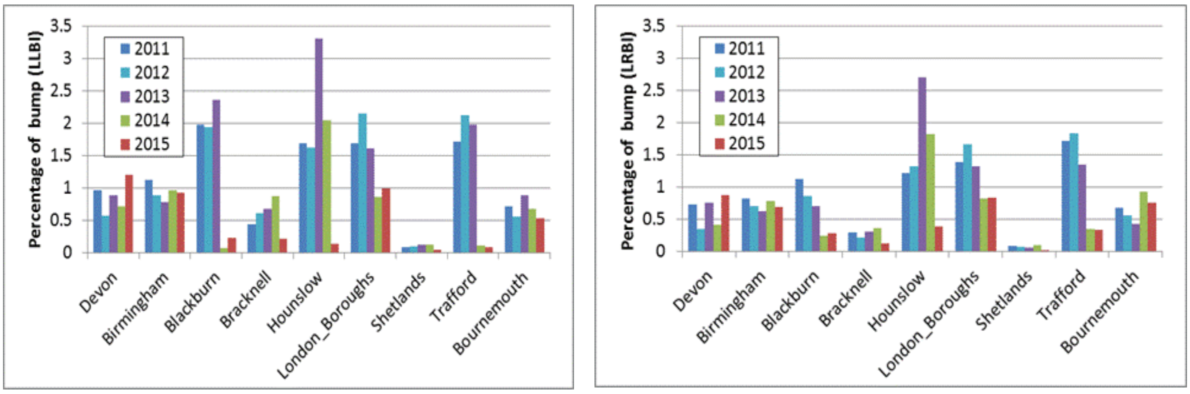



**Figure 13: Percentage of lengths containing a bump in the nearside (left) and the offside (right) for survey years 2011/12 to 2015/16**

A further investigation has been carried out on specific sites to determine why the Bump Measure is so inconsistent and whether it would be practical to update it to provide a more consistent measure (Section C.2). By comparing bump data with video images the investigation has shown that the bump measure does provide useful data on real bump-features, but it is inconsistent as to whether a bump gets reported or not, and the cause is not obvious. It may be due to sensitivity to driving line, since the measure is calculated from a very thin longitudinal measurement line, or may be due to the way that the parameter is calculated. It has not been possible, within the scope of the current project, to investigate this further.

*It is therefore suggested that a more robust measure may be achieved by considering the whole of the road shape, which would better model the bumps and would overcome issues with driving line. However, this would be a longer term development.*

# 7 Task 2: SCANNER Condition Parameters - Longer Term Development

We define *longer term developments* as enhancements that would probably require a development phase over the next 12-24 months followed by implementation. We have identified the following potential longer term developments from the results of the consistency work in Task 1, and the changes and improvements identified through the consultation.

## 7.1 Longer term development 1: Rutting

*The data delivery improvements for transverse profile recommended in Task 1 should be implemented via an update to the SCANNER specification, as discussed in Section 4.3.* As this will require updates to equipment and processing systems, we have classified this as a longer term development. However, if a specification revision is provided in 2017, trial data could commence delivery from the start of the 2018 survey.

*With the transition to high resolution transverse profile a replacement for the current rut measure should be considered. This enhanced measure could be reported alongside the current standard rutting for a number of years, until the new measure was deemed acceptable (i.e. it could be run in parallel).*

In Section 3, it was shown that introducing high resolution transverse profiles, a road marking profile and an algorithm that would eliminate measurements made on road markings and those lying outside of the lane being surveyed, would improve the consistency and accuracy of the rut depths calculated. Whilst the contractors are capable of providing the raw measurement data with the systems that they currently use, the rut depth calculating algorithms that they use may not be able to cope with this. It was also observed that the two fleets provide different levels of rutting on the same sites (within Accreditation tolerances). This is influenced by the different algorithms used by different contractors. Thus it may be beneficial for a single algorithm to be implemented for SCANNER.

The TRACS rutting algorithm was therefore assessed to determine its suitability for use for calculating rut depths on the local roads surveyed by SCANNER. It was found that, for principal roads and relatively wide non-principal roads, the TRACS algorithm identified the edge of the road well and also calculated more consistent and accurate rut depths. However, on narrower low class roads, it did not always place the simulated straight edge in a consistent or sensible position on the transverse profile. Further work would be needed to improve this, including:

- Prevention of straight edge being placed too close to the lane edge;
- Prevention of too much overlap between the straight edges used to calculate nearside and offside ruts;
- Prevention of straight edge for nearside being placed in offside of profile and vice versa;
- Reporting when the transverse profile is too narrow to calculate rut depths – where transverse variance would be a more appropriate parameter.

## 7.2 Longer term development 2 – Fretting Measure

*There is a desire for a measure of fretting. However, the three lines of RMST, currently provided by the SCANNER survey are not able to provide a good enough estimate to meet the users' needs. There is therefore a need to determine the measurements needed for this and to develop a fretting parameter.*

Prior to 2006, texture depth measurements made by laser systems in the UK were only reported as Sensor Measured Texture Depth (SMTD), which are generally reported at 10m intervals in the nearside wheeltrack. Alternative means of reporting the texture measurements made by current laser systems were also available, including the Mean Profile Depth (MPD) measure, which is widely used in Europe.

Research, carried out by TRL (Viner et al., 2006), determined that improved methods for detecting localised variability in texture were needed for B and C roads, due to the variability in texture that could be seen across the width of such roads. A method was demonstrated that used texture data collected across the lane width and combined information about the average level of texture depth, the overall variability and the difference between the centre of the lane and the wheel paths to assess the condition of the surface texture at a network level. This method was shown to be as good as a single measurement of texture depth in the nearside wheel path for identifying sections with deteriorating surface texture on a test dataset that included mainly roads with relatively high levels of surface texture. On roads with low surface texture the new method was expected to outperform the current nearside measurement. As a result of this work, it was recommended that the specification for SCANNER surveys included a measurement of transverse texture variability, in addition to the measurement of SMTD in the nearside wheelpath.

Due to the level of technology available on SCANNER vehicles at the time there was a need to restrict the technological demands for the measurement of transverse texture variability. Although a texture measure across the full lane width would ideally be provided to calculate the variability, it was practical to require the measurement of texture in only 3 lines. Unfortunately, experience has shown that fretting is a more important defect to road engineers than basic variability, and this cannot be determined from the three measurements.

Developments in equipment now offer the potential for the required full lane width texture data, and this is a requirement of TRACS surveys from 2017. Fretting parameters are calculated from multiple line measurements TRACS (Benbow et al., 2011). And this may be achievable on local roads if SCANNER were to deliver the required texture data, which we believe to be achievable using current SCANNER equipment. For example Figure 14 and Figure 15 show the forward facing image of a surface defect and the corresponding high resolution multiple line texture (RMST) data respectively. As can be seen from Figure 15, the RMST data allows for the shape and detail of the defect surrounding the patches to be clearly identified.

Similarly, Figure 16 and Figure 17 show the forward facing image and RMST plot for a section of fretting on a local road. Again the high resolution RMST data allows for the transverse and longitudinal extent of the fretting to be identified.

**Figure 14: The forward facing image of a pavement defect surrounding a patch, seen on the SRR2 extension route**

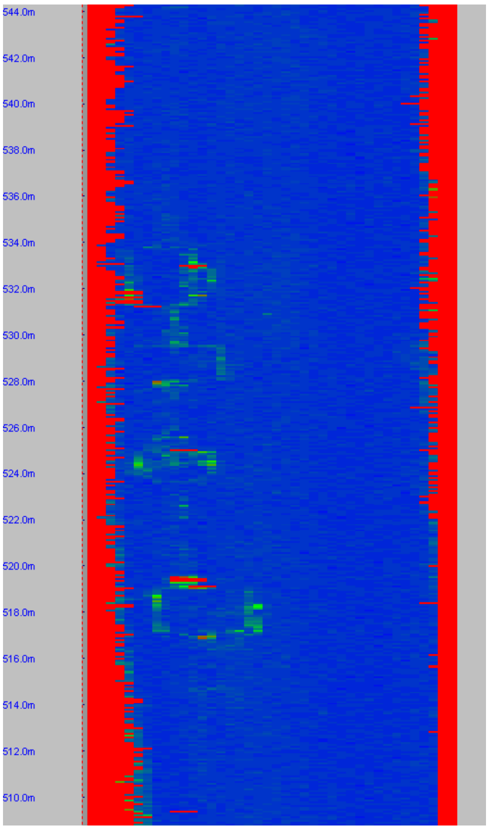



**Figure 15: The RMST plot of the defect shown in Figure 14, which clearly shows higher RMST values (using high resolution transverse profile data to provide 40 RMST values across the road width)**

**Figure 16: The forward facing image of fretting, seen on the SRR2 extension route**

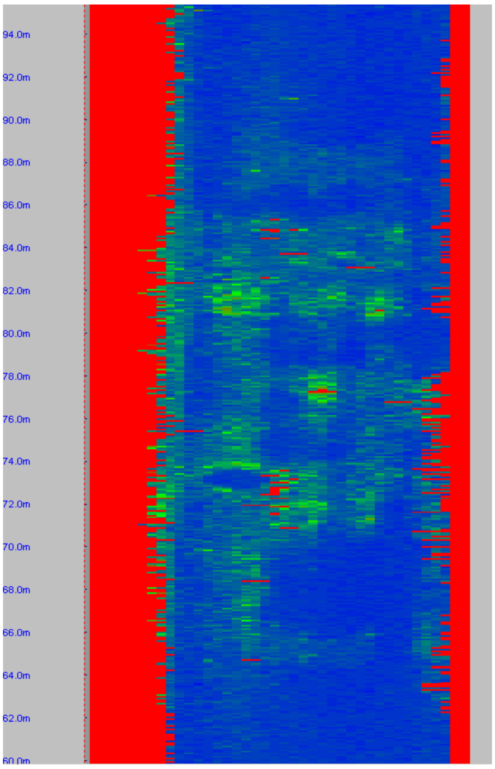



**Figure 17: The RMST plot of the fretted area shown in Figure 16, which clearly shows higher RMST values (using high resolution transverse profile data to provide 40 RMST values across the road width)**

## 7.3 Longer term development 3 – bump/pothole measure

*If the current Bump Measure cannot provide a network level indicator of the extent to which the network is affected by bump causing features, or of potholes (see quick win above), there will be a need to develop a different parameter to achieve this.*

It has been shown above that the limitations of the bump measure are likely to be fundamentally linked to the two measurement lines it is able to provide. As with full lane width texture data, we believe that it should also be possible to obtain full width longitudinal profile measurements using current SCANNER equipment. It may be possible to calculate a pothole/lane width bump measure using this data. Thus a lane width measure should be achievable on local roads if SCANNER were to deliver the required profile data.

For example Figure 18 and Figure 19 show the downward facing image of surface defects, the 3D profile (extracted from HARRIS2's high resolution transverse profile measurement system) and the corresponding results from a 3D version of the current Bump Measure. As can be seen, the 3D Bump Measure data allows for the shape and detail of the defects to be clearly identified. Note that the features shown on the 3D profile plot appear more stretched on the right hand side, due to the way that this data has been plotted.

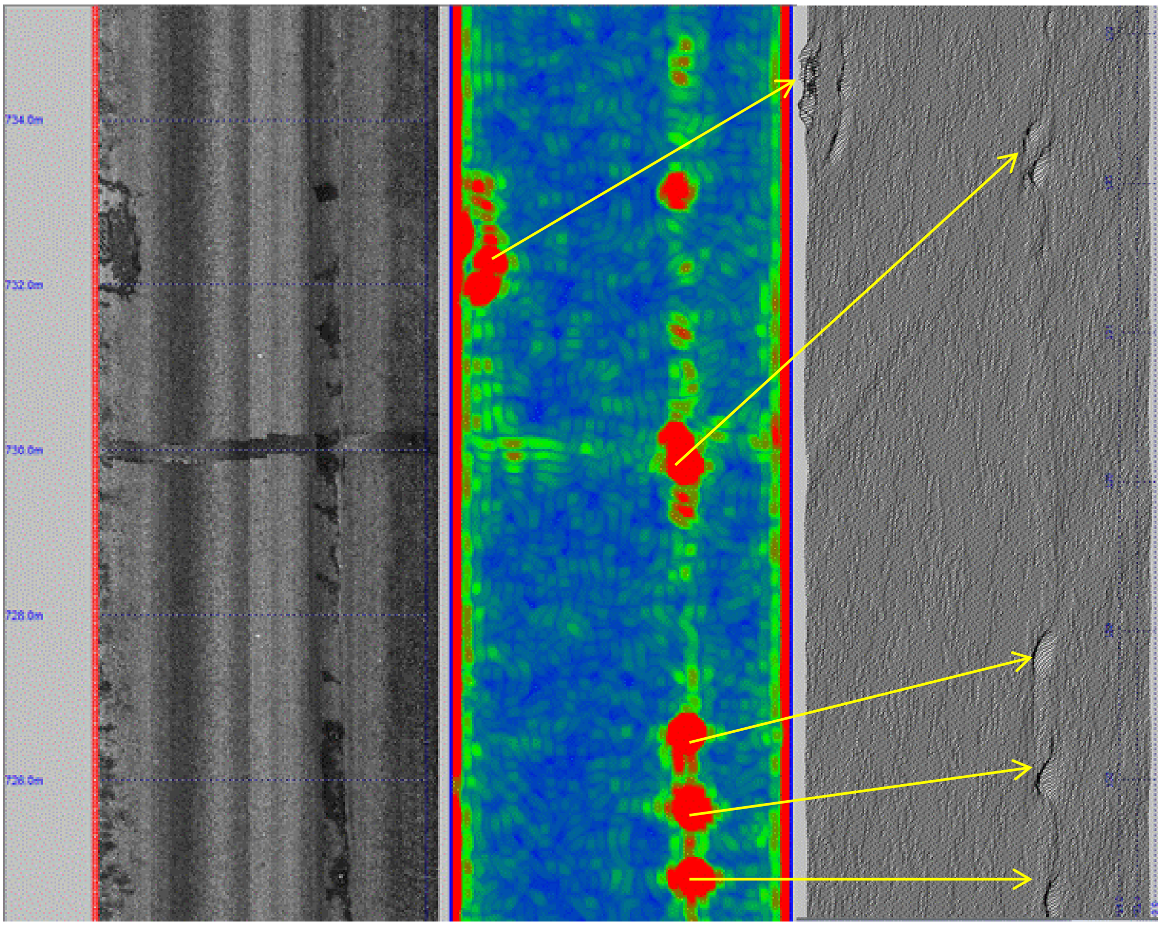


**Figure 18: Downward facing image showing failing patch around a gully and several dips (left), results of applying 3D Bump Measure to the 3D profile data (right)**

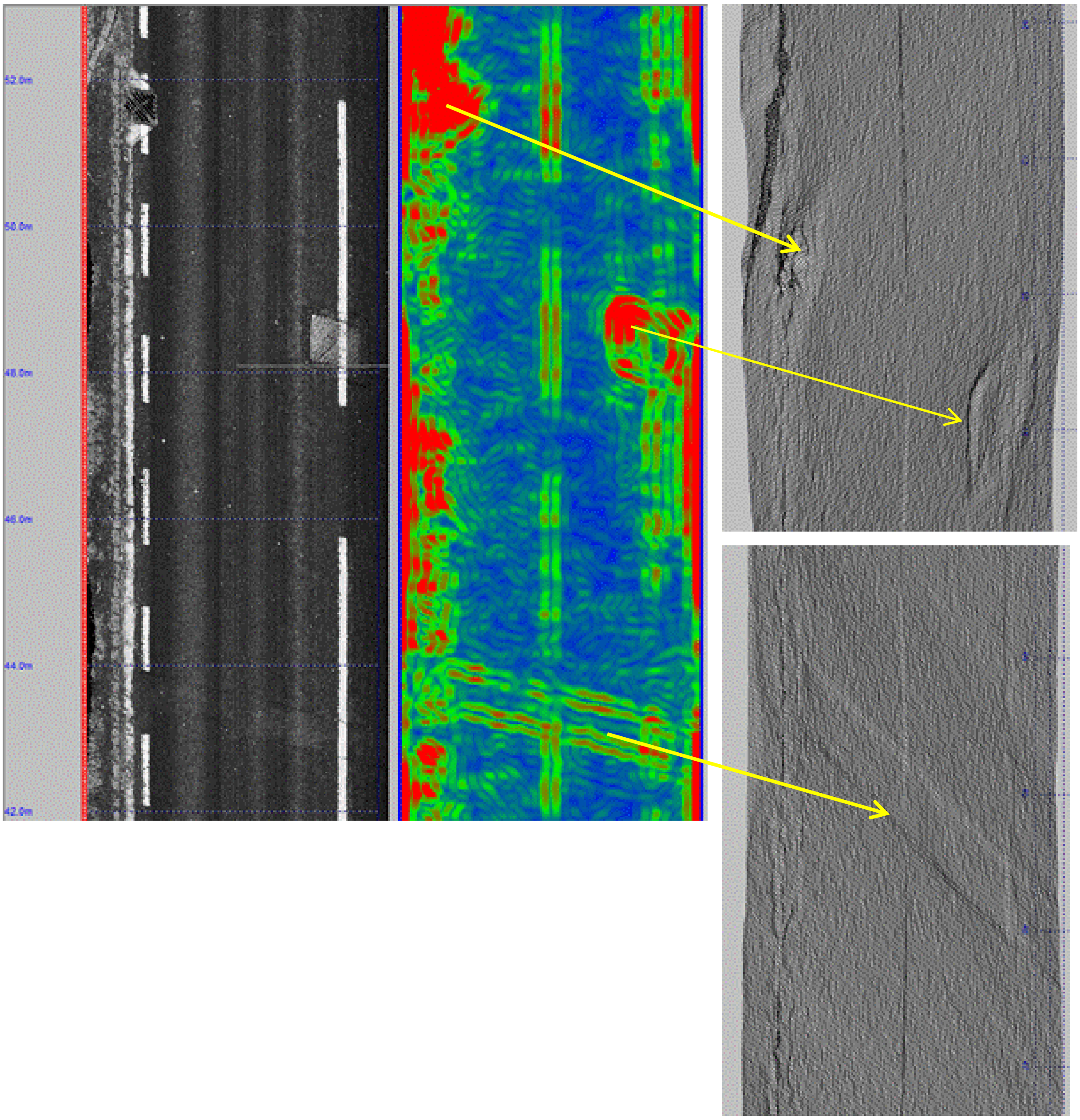


**Figure 19: Downward facing image showing failing patch around a gully, a sunken manhole cover and a sunken transverse trench (left), results of applying 3D Bump Measure to the 3D profile data (right)**

## 7.4 Longer term development 4 - Training

*There is a need to develop an education strategy for use of SCANNER data (parameters), which should be developed alongside the education recommended by Task 3 (Spong & Cartwright, 2017).*

The purpose of the strategy will be to develop local authority confidence and expertise in the use of SCANNER data. It is envisaged that delivery is likely to be via high-quality multimedia education materials so that the courses are inclusive and accessible regardless of location, time constraints or other local limitations.

This will enable full use to be made of SCANNER data by developing confidence and expertise throughout the industry. It will also provide a platform for new developments to be disseminated in the future. The education strategy is a vital ingredient in helping the industry to gain the greatest benefit from other improvements to asset management by ensuring that they reach as wide an audience as possible and are implemented to maximum effect.

# 8 Summary and Recommendations

## 8.1 Consistency of SCANNER

The first task of this project has investigated the consistency of SCANNER. The focus has been on rutting and cracking, which were identified as key consistency concerns in the SCANNER Development Group review.

The consistency of the cracking data has a significant effect on the year to year consistency of network level reporting. Hence cracking has been observed to be the main cause of the large inconsistencies seen in the QA audit process. Although the assessment has suggested that there might be differences between the level consistency of cracking on rural and urban roads, this was not strongly shown in individual LHAs. However, cracking data collected during the winter months *was* observed to be less consistent than data collected during the summer. Therefore it is recommended that a winter shutdown is implemented, which will require discussion with the survey industry.

There is currently no method to check that the fleet is consistent, and the repeatability test is also weak. Therefore the work has developed enhancements to the Accreditation process for cracking to improve these tests. As a result it is recommended that a new test for repeatability, as devised within this project, is implemented immediately. A new test for fleet consistency has also been devised within this project. This is a more complex test, that will require experience to understand its effect on the current SCANNER fleet. It is therefore recommended that this test is implemented now and trialled over the next 12 months, to allow SCANNER contractors time to develop an action plan to improve any devices found to be inconsistent. It would become a formal requirement at the end of the trial.

Rutting is generally considered a reasonably consistent parameter. However, whilst inconsistencies tend to be small, they can become significant when combined with other parameters, to influence the RCI. Overall the fleet has become more consistent in the last few years. However, there is a noticeable difference between the fleets of the two current contractors, with an average difference in rut depth of 1.7mm being reported.

Possible routes to improve the consistency of rutting have been investigated. These have included development in both the collection and the processing technologies. It has been shown that using higher resolution systems, with wider measurement width, combined with road marking removal, could provide more accurate and repeatable data. Using a centrally defined and controlled rut algorithm could also reduce the fleet inconsistency. As all SCANNER contractors now employ high resolution systems (and sample their data down for current SCANNER delivery), it would be feasible to increase the performance requirements defined in the SCANNER specification to require the delivery of this data. In addition, the TRACS rutting algorithm has been trialled and found that, subject to improvements to the automatic edge detection algorithm and the placement of the straight edge, it should be able to provide good performance on the lower classes of roads. Therefore it is recommended that implementation of these updates to the SCANNER requirements should be considered.

## 8.2 SCANNER condition parameters

SCANNER delivers more than 20 parameters but only a few are used to calculate the Road Condition Indicator (RCI). Also few LHAs make use of the enhanced parameters provided in the 2007 research. Conversely, the survey does not provide all the condition parameters that are considered to be important by LHAs. Better value could be obtained from SCANNER if the parameters were optimised to reflect LHA needs. LHAs and PMS providers have been consulted to identify potential revisions/enhancements to the SCANNER condition parameters, or potential new parameters that could be included in a future SCANNER survey. Several observations and recommendations resulted from this consultation, which have been used to identify a number of potential **quick wins** and **longer term developments.**

### *8.2.1 Quick wins*

**Quick wins** are enhancements that could be implemented in the next 12 months. These include:

**Cracking**

- The consistency improvements recommended in Task 1 should be implemented as soon as practical;
- Of the delivered cracking data, value is being drawn from Whole Carriageway Cracking and Wheeltrack Cracking only. The remaining surface deterioration parameters are not required in the delivered data.

**Ride Quality**

- Use is only being made of one of the two profile (roughness) parameters. Therefore LPV should be phased out and replaced with eLPV. This will deliver a more stable and accurate RCI, and will reduce the adverse effect of geometry on the data;
- The measurement of roughness is currently failing to report defects present in the offside wheelpath. The measurements from both wheelpaths should be included in the RCI calculation, to provide a more robust assessment of ride quality.

### *8.2.2 Longer term developments*

**Longer term developments** are enhancements that would require a development phase over the next 12-24 months, followed by implementation. These Include:

**Rutting**

- The improvements to transverse profile recommended in Task 1 should be implemented as soon as practicable. Delivery of this wider, higher resolution profile will contribute to improvements in rutting accuracy and repeatability. With the transition to high resolution transverse profile a replacement for the current rut measure should also be considered. This could be via a single rut algorithm across all SCANNER devices, which would minimise the differences arising from the use of different algorithms by different contractors. The new rutting algorithm could be trialled alongside the current rutting, until the new measure was deemed acceptable;
- Even with an updated rut algorithm, rut depth is sometimes an inappropriate measure to use on narrow roads (e.g. U roads). Transverse variance would be a more appropriate parameter on these roads. The use of this parameter should be considered further.

**Fretting**

- There is a clear call from LHAs for a measure of fretting. The current SCANNER texture variability provides a poor proxy for this.
- The use of multiple line texture measurements, extracted from high resolution transverse profile data, shows promise for the identification of fretting. It is recommended that a method be developed to obtain fretting from this data, to hence deliver a fretting parameter.

**Bump/pothole measure**

- There has been a strong request for potholes to be included in SCANNER. The current SCANNER Bump Measure does not provide a reliable network level indicator of the extent to which the network is affected by such features.
- High resolution transverse profile data could be adopted to provide full lane width longitudinal profile data, from which a more reliable bump/pothole measure could be obtained. Development of this parameter is recommended.

**Training**

- There is a need to develop an education strategy for use of SCANNER data (parameters). This could be developed alongside the education recommended by Task 3, to hence include the survey, its measurements and the uses of the data (RCI/UKPMS).
- The purpose of the strategy will be to develop local authority confidence and expertise in the use of SCANNER data. It is envisaged that delivery is likely to be via high-quality multimedia education materials so that the courses are inclusive and accessible regardless of location, time constraints or other local limitations.

# 9 Implementation Plans

This section presents proposed implementation plans for the delivery of the quick wins and longer term developments given in Sections 6 and 7.

## 9.1 Quick Win 1: Cracking

| | |
|---|---|
| **Objective** | **Add tests for fleet consistency and device repeatability to the existing Accreditation procedure.** |
| **Purpose** | It has been shown that devices within the SCANNER fleet are not consistent with each other and were often not consistent with themselves. This aims to overcome this problem through the introduction of focussed tests to drive improvements in the fleet. |
| **Benefits** | Devices showing poor performance will be better identified and can be removed from the surveying fleet. Those devices with better performance will be encouraged to continue to provide the required level of consistency throughout their lifetime. Thus, it is expected that these changes this will improve the overall consistency of the cracking measure. |
| **Implementation Plan** | There already an outline requirement for repeatability within the Accreditation tests. The proposed repeatability tests are a formalisation of this process. Test for repeatability should be implemented within Accreditation as soon as possible (within the 2017 tests). This will require<br>• The SCANNER specification to be updated and published<br>• The tests to be implemented within the Accreditation tests.<br><br>Testing of fleet consistency has not previously been implemented. Including these in Accreditation may result in a number of contractor devices failing. The fleet consistency test should be introduced as soon as possible but not enforced within Accreditation for at least one year. This will allow a trial period for the method, to identify any issues that need to be ironed out before full implementation, and will provide contractors time to determine which of their devices might be inconsistent with the fleet and to develop an action plan to improve this. This will require:<br>• Draft amendments to the current SCANNER specification to be delivered<br>• The tests to be implemented within the Accreditation tests (but not enforced), to the draft amendments<br>• Discussion with the contractors to ensure that the process will work when implemented fully.<br>• After the trial period, the amendments to the SCANNER specification will need to be formally implemented and published and the tests enforced within the Accreditation tests. |
| **Risks, Issues and Dependencies** | If these changes are not implemented, the consistency of the cracking parameter will not improve. This will reduce the use of cracking data further, potentially undermine the users' perception of data quality (for all parameters) and thus reduce the value for money that the SCANNER survey currently provides. |

## 9.2 Quick Win 2: Ride Quality

| | |
|---|---|
| **Objective** | **Phase out the use of the LPV ride quality measure and replace it with the alternative eLPV measure**.<br><br>**Addition of offside ride quality measurements in RCI calculation** |
| **Purpose** | Two ride quality parameters are currently provided by the SCANNER survey: LPV and eLPV. Since it is used in the RCI calculation, only LPV is used currently by stakeholders. It has been shown that LPV is adversely affected by road geometry, and this led to its replacement, by eLPV, in the TRACS survey in June 2004. This work would implement a phased introduction and transfer to eLPV from LPV.<br><br>The longitudinal profile is measured in both the nearside and offside wheelpath. However only LPV, calculated from nearside measurements, is included in the RCI |
| **Benefits** | The eLPV parameter is much less affected by road geometry than LPV, so provides a more reliable and consistent measure on lengths with varying extremeness of road geometry (hills, bends etc.). Thus it is a more robust measure. Using eLPV in place of LPV would align SCANNER with TRACS.<br><br>Including the offside measurement in the RCI calculation would result in the RCI better reflecting user opinion. |
| **Implementation Plan** | Implement via UKPMS developers:<br>• For the next 5 years, continue to provide LPV and eLPV within HMDIF.<br>• UKPMS to provide additional RCI and national indicators calculated using eLPV, instead of LPV. This will require the development of a rules and parameters set for eLPV within UKPMS. As with rutting, using 2 lines of eLPV would use max(NS eLPV, OS eLPV) in the RCI calculation<br>• Educate the users on these two indicator sets and the likely difference that will be seen between them<br>• After 5 years, stop providing LPV and calculate the RCI with 2 line eLPV. This will require a change to the SCANNER specification. |
| **Risks, Issues and Dependencies** | Continued use of LPV, a measure that is known to provide high values, even when the ride quality is good, undermines the users' trust in the data, leading to diminished use of the data, resulting in poor value for money from the survey. Authorities with networks that have varying geometry will continue to have higher levels of RCI reported on lengths where there is no reason to improve the ride quality, leading to inaccurate assessment of network condition.<br><br>Only including the measurements of ride quality from the nearside wheelpath will result in the lack of identification of many lengths with poor ride quality. This should be implemented at the same time as replacing LPV with eLPV. |

## 9.3 Quick Win 3 – Bump measure

It has been determined that this will require longer term development – see below.

## 9.4 Longer term development 1a: Rutting - High resolution transverse profile data

(This would be run in parallel with longer term development 1b)

| | |
|---|---|
| **Objective** | **Improve the measurement of transverse profile in SCANNER by better utilising the capability of current equipment, to deliver high resolution transverse profiles and also a road marking profile** |
| **Purpose** | The inconsistency in the rut measurements for the vehicles in the current fleet have been shown to be significant and likely to arise from driving line, resolution, survey width and the rut algorithm used.<br><br>The purpose of this work would be to implement improvements to the transverse profile measurement to support resolving this issue. |
| **Benefits** | There has been a step change in the measurement systems since SCANNER was developed. The contractors now use high resolution systems and "dumb down" the data. This development would better utilise these systems.<br><br>Using higher resolution transverse profiles and road marking removal should provide data that will enable improvements to the calculation of rut depths, so that they are more accurate and more repeatable. It is also expected to improve the fleet consistency seen for rutting. |
| **Implementation Plan** | Implement via changes to the SCANNER specification<br><br>• Amendment of the SCANNER specification to require delivery of<br>    o High resolution transverse profile<br>    o Road marking profile<br>    o Standard Rutting from high resolution transverse profile<br>• Amendment to include testing/accreditation of road marking profile. |
| **Risks, Issues and Dependencies** | If these changes are not implemented, this will result in the consistency of rutting remaining the same. It will also results in the contractor's systems not being fully utilised, thus offering poor value for money for the survey.<br><br>Implementation of a new rutting algorithm should be carried out to achieve the full benefits of this. |

## 9.5 Longer term development 1b: Rutting – rut algorithm

(This would be run in parallel with longer term development 1a)

| | |
|---|---|
| **Objective** | **Deliver enhancements to the SCANNER rutting algorithm** |
| **Purpose** | The inconsistency in the rut measurements for the vehicles in the current fleet have been shown to be significant and likely to arise from driving line, resolution, survey width and the rut algorithm used.<br>The purpose of this work would be to implement improvements to the processing of the transverse profile measurement to deliver a single rut algorithm that could be used by any system in the fleet system. |
| **Benefits** | Unifies calculation of rutting: Any differences seen between contractors will then be narrowed down to the measurement of transverse profile, not the rut calculation. It is expected to improve the accuracy and repeatability of rut depth reporting across different authorities. |
| **Implementation Plan** | The work would build on existing rutting algorithms (e.g. the TRACS rutting algorithm) to improve its performance (e.g. straight edge placement) and in particular how the algorithm can be optimised for lower classes of road, where a verge is present.<br>Once this development work is finished, the algorithm definition will be added to the specification and published.<br>A phased implementation would then deliver the high-resolution rutting data alongside the current rutting measure to understand/manage any effect of the new algorithm on the RCI, national reporting etc. |
| **Risks, Issues and Dependencies** | If these changes are not implemented, this will result in the consistency of rutting remaining the same.<br>Implementation of high resolution transverse profile and road marking profile is required for this. |

## 9.6 Longer term development 2 - Fretting Measure

| | |
|---|---|
| **Objective** | **Develop a SCANNER fretting parameter** |
| **Purpose** | The current SCANNER survey does not provide a measure of the fretting present on the network but this was identified as an important defect by the stakeholders.<br><br>The purpose of this development would be to determine the data requirements for the measurement of fretting from SCANNER data, to develop the algorithms to identify fretting and to deliver them for implementation on the network. |
| **Benefits** | This will provide a measure that is important to stakeholders, but that is not currently provided by the survey. It will provide better value for money from the survey, and data that better meets the needs of the data users. It will enable improved asset condition assessment and will support maintenance identification, planning and design. |
| **Implementation Plan** | Review the use of current high resolution transverse profile measurement systems to confirm how/what they will be able to provide to support the measurement of fretting (e.g. multiple lines of RMST data). Hence define the data delivery requirements for SCANNER survey vehicles. Note: this will be based on existing capability – the aim will be to achieve a fretting measure that can use current SCANNER capability.<br><br>Develop an algorithm to utilise this data provide a measure of fretting (likely to be based on methodology developed for strategic roads, and also drawing on technical developments elsewhere in Europe).<br><br>Define the methodology, via updates to the SCANNER specification, and work with SCANNER contractors to implement this. |
| **Risks, Issues and Dependencies** | If these developments are not made SCANNER will continue to deliver an “incomplete” dataset. LHAs will continue to rely on other data sources, that may be less robust, less objective and that incur additional survey costs.<br><br>The developments will benefit from working alongside contractors to ensure a cost effective, practical and implementable result. |

## 9.7 Longer term development 3 – Bump/pothole Measure

| | |
|---|---|
| **Objective** | **Develop a more robust bump/pothole measure to provide a network level indication of the number of lengths affected by features that would cause user discomfort.** |
| **Purpose** | Stakeholders have requested a measure of potholes on the network. The current Bump Measure was an attempt at this but has been found to be inconsistent.<br><br>This work will draw on high resolution profile data to overcome the consistency issues associated with the bump measure, and provide a better measure of the lengths affected by bump-like features such as failed patches and potholes. |
| **Benefits** | Will provide better value for money from the SCANNER survey, due to existing measurements (longitudinal and transverse profile) being better utilised to provide additional parameters that the users have expressed a need for. |
| **Implementation Plan** | Investigate the use of high resolution transverse profile data, combined with longitudinal profile data, to provide a more robust measure of bumps and potholes that is not so affected by driving line as the current measure:<br>• Review current high resolution transverse profile measurement systems to confirm how/what they will be able to provide to support the measurement of bumps/potholes<br>• Development of new characterisation method for this data<br>• Collection of reference data, to verify method<br>• Produce algorithm definition for new method<br>• Update SCANNER specification. |
| **Risks, Issues and Dependencies** | If these developments are not made SCANNER will continue to deliver an “incomplete” dataset. LHAs will continue to rely on other data sources, that may be less robust, less objective and that incur additional survey costs.<br><br>The development is reliant on the provision of high resolution transverse profile with suitable capability to enable the detection of features “along the road”. They will also benefit from working alongside contractors to ensure a cost effective, practical and implementable result.<br><br>Note that due to the frequency of surveys, a SCANNER pothole measure is not appropriate for rapid detection of these defects (which would still be identified in routine safety inspections). However, it can provide a measure (snapshot indication) of the length affected by these features, which is valuable for asset management and local/national performance tracking. |

## 9.8 Longer term development 4 - Training

| | |
|---|---|
| **Objective** | Develop an education strategy for use of SCANNER data within UKPMS |
| **Purpose** | This task is extensive and moves beyond communication and awareness to a full-blown education strategy. The purpose of the strategy will be to develop local authority confidence and expertise in the use of SCANNER data. It is envisaged that delivery is likely to be via high-quality multimedia education materials so that the courses are inclusive and accessible regardless of location, time constraints or other local limitations. |
| **Benefits** | This will enable full use to be made of SCANNER data by developing confidence and expertise throughout the industry. It will also provide a platform for new developments to be disseminated in the future. The education strategy is a vital ingredient in helping the industry to gain the greatest benefit from other improvements to asset management by ensuring that they reach as wide an audience as possible and are implemented to maximum effect. |
| **Implementation Plan** | 1. Define scope of task<br>2. Develop ongoing education strategy, including potential multi-media channels<br>3. Produce education framework (dependent on previous steps)<br>4. Deliver training materials to initiate approach (dependent on previous steps) |
| **Risks, Issues and Dependencies** | This is linked to the education task recommended by Task 3 (Spong & Cartwright, 2017) and should be performed alongside that task. |

# 10 Acknowledgements

The authors are grateful to the input provided by the following:

**SCANNER Development Group**

Amanda Richards, RCMG, Surrey County Council

John Hunter, Norfolk County Council

Graeme Ferguson, SCOTS, SRMCS, Perth and Kinross Council

Stephen Sherwood, Bridgend County Borough Council

Steve Batchelor, Highway Surveyors

Ian Butler, WDM

Dr Manish Jethwa, Yotta

Steve McQueen, WDM

Dave Rieley, Fugro Aperio

Ashley Singh, DfT

John Wilkins, DfT

Dr Dean Wright, Yotta

**Local Authority Consultees**

Marc Tite, Bristol City Council

Chris Nelson, Carmarthenshire County Council

Eric Opoku Ohemeng, Colas (Portsmouth City Council)

Andy Stevenson, Cornwall County Council

David North, Cumbria County Council

Aileen Woods, Dumfries & Galloway Council

David Prudence, Essex County Council

Sara Grindley & Nainesh Patel, Kier (Northamptonshire County Council)

Simon Jones, Metis Consultants (LB Kingston)

Haz Chudasama, Leicester City Council

John Hunter, Norfolk County Council

John Scougall, North Lanarkshire Council

Joe Lawson, DRD Northern Ireland

James Dance, Oxfordshire County Council

Graeme Ferguson, SCOTS

Crawford Lindsay, South Lanarkshire Council

Peter Burnham, Worcestershire County Council

# 11 References

**Agresti A and B.A. Coull B.A. (1998)**: "Approximate is better than 'exact' for interval estimation of binomial proportions". The American Statistician, Vol. 52, 1998.

**Benbow E and A. Wright (2008)**: "Assessment of the Consistency of SCANNER Parameters, RCI and BVPI". TRL Published Project Report PPR 321.

**Benbow E, K Nesnas and A Wright (2006):** "Shape (surface form) of local roads". TRL Published Project Report PPR 131.

**Benbow E, S McRobbie, N Dhillon, K Nesnas, D Wright and A Wright (2011):** "Development of the requirements for the TRACS3 survey". TRL Client Project Report CPR1554.

**Chris Britton Consultancy (CBC) (2006)**: "Consistency of TTS on Local Roads". January 2006.

**McRobbie S, L Walter, C Read, H Viner & A Wright (2007):** "Developing SCANNER Road Condition Indicator parameter thresholds and weightings". TRL Published Project Report PPR 199

**McRobbie S, K Nesnas, S Britain & M Harrington (2017)**: "Improved scheme identification and prioritisation by automated detection of change". TRL Published Project Report PPR 814

**SCANNER survey for Local Roads User Guide and Specification - Volume 5 (2009):** Technical Requirements for SCANNER Survey Parameters and Accreditation, http://www.ukroadsliaisongroup.org/en/guidance/scanner-guidance-documents.cfm

**Spong C and R Cartwright (2017):** "Development of SCANNER and UKPMS. Task 3: Appropriateness of the SCANNER RCI". TRL Project Report PPR817

**Viner H, P Abbott, A Dunford, N Dhillon, L Parsley and C Read (2006):** "Surface Texture Measurement on Local Roads". TRL report PPR148.

**Wright D, V Baltazart, N Elsworth, R Hamrouche, J Krarup, M L Antunes, S McRobbie, V Merecos, B Meitei and T Saarenketo (2014):** "Monitoring structural and surface conditions". Deliverable 4.3 of the TRIMM project http://trimm.fehrl.org/?m=3&id_directory=7539

# Appendix A Consistency of SCANNER cracking data (Task 1)

## A.1 Task 1: Consistency of SCANNER cracking data

### *A.1.1 Effect of cracking on RCI*

The RCI combines the rutting, roughness (LPV), texture and cracking data to obtain an overall score that is used to report the condition of each 10m length. The UKPMS rules and parameters define the thresholds and weightings for calculating the RCI. These rules apply a weighting of only 0.6 to cracking. This means that cracking has less influence than other parameters and would suggest that inconsistency in the cracking data shouldn't have a large effect on the RCI. However, experience gained from network QA shows cracking often has the most significant effect on the RCI. During the QA process, SCANNER data from subsequent years is compared, to determine the consistency between the years. The QA process includes an Audit Indicator, which calculates the percentage of lengths in that year's data having an RCI≥100. This is compared with the previous year's Audit Indicator. As there is expected to be some stability in the data, Authorities where there are significant changes in the Audit Indicator are investigated to determine if the change is associated with poor data quality.

The 2015 QA audits identified Bournemouth, Glasgow, and Clackmannanshire as examples of Authorities having potential issues with consistency, with Bournemouth seeing an increase in their Audit Indicator from 3.1% in 2013/14 to 5.5% in 2015/16. (Since the requirement is for condition measurements to be made on A roads every 2 years and on B and C roads every 4 years, comparing data from subsequent years can lead to data being compared from very different parts of the network. Therefore, data from two years apart is often compared to ensure that data from sufficient common lengths is contained in the two datasets). Further assessment showed that the cracking for this authority was very inconsistent for the two years (Figure 20).

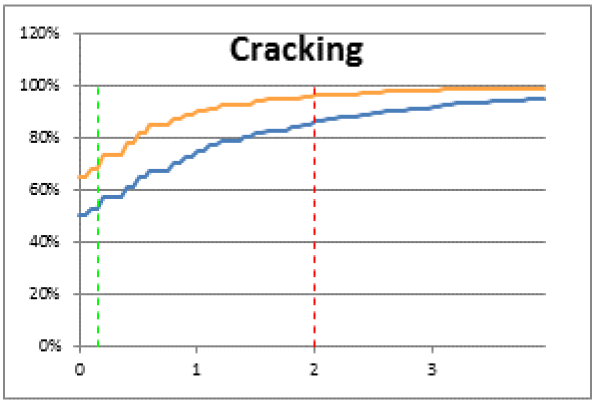



**Figure 20: Cumulative frequency distribution for cracking on the Bournemouth network: Blue line is 2015/16 data, orange is 2013/15 data**

To answer the question over how much the cracking contributes to changes in the overall indicator, the cracking data from 2013/14 was substituted into the RCI calculation for 2015/16, in order to determine the effect of cracking on the RCI. This resulted in the Audit indicator for 2015/16 being reduced from 5.5% to 2.8% and the percentage of Green and Amber lengths being more similar, as shown in Table 6.

**Table 6: Effect of cracking data consistency on RCI**

| Year | Red (RCI ≥ 100) | Amber (20<RCI<100) | Green (RCI ≤ 20) |
|---|---|---|---|
| 2015/16 | 5.5% | 28.8% | 65.7% |
| 2013/14 | 3.1% | 21.0% | 76.0% |
| 2015/16 with 2013/14 cracking data | 2.8% | 20.5% | 76.7% |

This suggests that cracking is the main cause for the large inconsistency between the two years. The situation was similar for Glasgow and Clackmannanshire. Thus we can conclude that inconsistency in cracking data can and does have a significant effect on network level reporting.

### *A.1.2 Road environment*

To assess the variability in cracking by road environment, 6 years' data was collated from the national SCANNER database of SCANNER survey data. The average cracking value was calculated, for each year, for

- All road types
- All rural roads
- All urban roads
- All principal roads (i.e. A class)
- All non-principal roads (i.e. B & C class), resulting in significant lengths of network data being considered (Table 7).

**Table 7: Lengths of network data used to calculate cracking average**

| Year | 2010/11 | 2011/12 | 2012/13 | 2013/14 | 2014/15 | 2015/16 |
|---|---|---|---|---|---|---|
| All data (km) | 133,674 | 111,646 | 133,961 | 136,439 | 139,139 | 125,705 |
| Rural | 78% | 77% | 79% | 80% | 80% | 80% |
| Urban | 22% | 23% | 21% | 20% | 20% | 20% |
| Principal | 39% | 38% | 37% | 38% | 36% | 38% |
| Non-principal | 61% | 62% | 63% | 62% | 64% | 62% |

The average cracking values over the last six years are plotted in Figure 21. For a consistent measure, we would expect that the average value would remain similar from year to year and thus the graphs would consist of fairly flat lines. However, it can be seen from Figure 21 that this is not the case, particularly for urban roads, which have a much larger variability between years than rural roads.

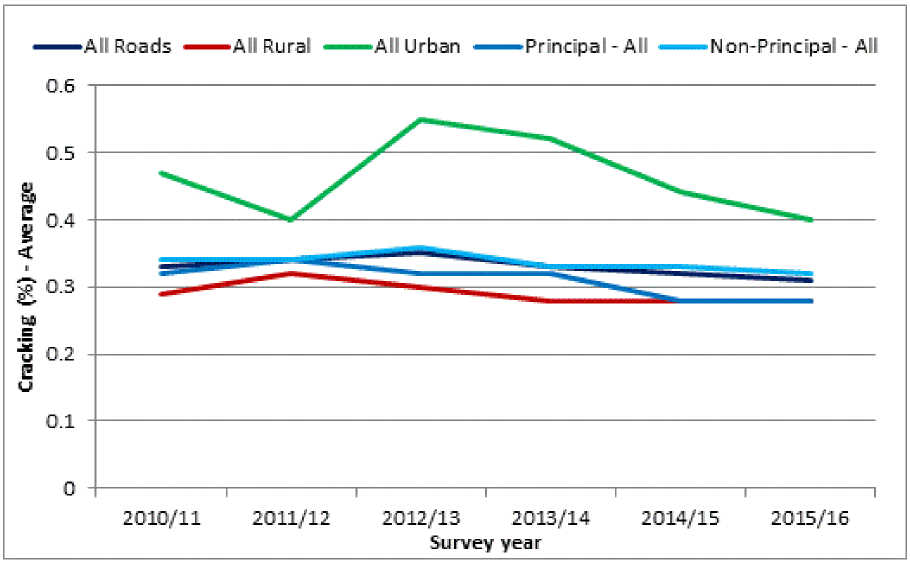



**Figure 21: Average cracking value on whole network for different road types**

It is possible that the average values have been affected by a small number of very large values (outliers) and thus the distribution of values was also inspected, to determine if such outliers were causing the inconsistency. As can be seen from Figure 22, this does not appear to be the case when all roads are considered. (Note that the distribution is not smooth due to the choice of bins for the distribution and also the discrete nature of the cracking parameter (it is always a multiple of the percentage of road surface covered by a single grid square)).

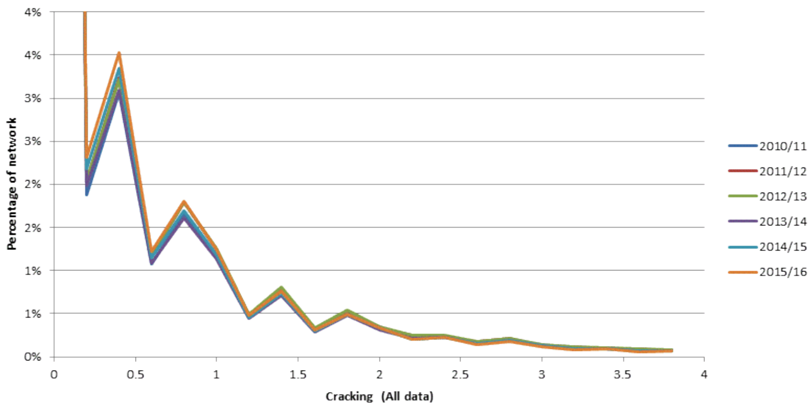



**Figure 22: Distribution of cracking values on all roads**

The situation is the same if just urban, rural, principal or non-principal roads are considered (Figure 23). As with the average values, the cracking looks more inconsistent on urban roads than any other type of road. Thus the inconsistency seen in the average values is caused by general inconsistency in the data, not by a small number of spikes (outliers).

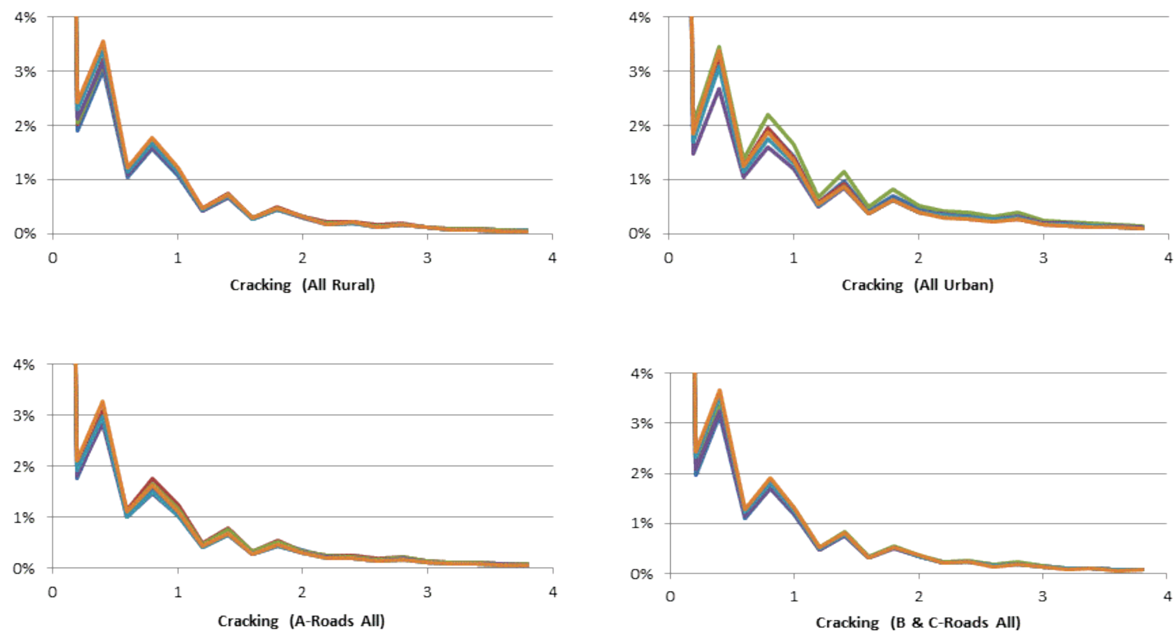



**Figure 23: Cracking distributions split by road type**

It is thought that the difference between consistency on urban roads and rural roads may be due to urban roads containing a higher frequency and larger range of features (reinstatements, road markings etc.) than rural roads, and thus false positives may provide the explanation for some of this inconsistency. However, neither raw data nor reference data was available to assess this result.

Since it has been seen that cracking from urban lengths may be a major source of inconsistency, it was thought that removing the urban roads from the Audit Indicator calculation might improve the consistency. To investigate this, data was used from an LHA, identified as having issues with RCI and cracking consistency by the QA Audit process: Clackmannanshire contains lengths of both rural (79%) and urban roads, as shown in Table 8.

**Table 8: Surveyed lengths in Clackmannanshire**

| Length | 2015 | 2014 | 2013 |
|---|---|---|---|
| All | 75.04km | 80.95km | 74.9km |
| Rural only | 59.86km | 64.79km | 59.77km |

Only considering rural roads from the Clackmannanshire network does improve the consistency of cracking slightly, in that the bias for cracking is slightly smaller (Table 9). It is still larger in size than the lower threshold use for RCI calculation though (0.15). However, the effect on the RCI, of removing these roads, is not significant. Thus it can be concluded that there is significant inconsistency in the cracking data for the rural roads on this network too.

Due to the low number of local authorities which contain reasonable amounts of both rural and urban roads, it was not possible to carry out further tests on other authority data. However, the results from Clackmannanshire would suggest that whilst the overall network assessment indicates a possible difference between rural and urban, this is not strongly shown in individual LHAs.

**Table 9: Consistency of Cracking in Clackmannanshire when all roads are considered and when only rural roads are considered.**

| | Year of data | Cracking bias from previous year | Red | Amber | Green |
|---|---|---|---|---|---|
| All roads | 2015/16 | -0.21 | 3.0% | 30.4% | 66.6% |
| | 2014/15 | | 10.8% | 36.0% | 53.2% |
| | 2015/16 | -0.26 | 3.0% | 30.4% | 66.6% |
| | 2013/14 | | 11.9% | 38.7% | 49.4% |
| Rural roads only | 2015/16 | -0.19 | 3.1% | 30.5% | 66.5% |
| | 2014/15 | | 10.0% | 35.8% | 54.1% |
| | 2015/16 | -0.20 | 3.1% | 30.5% | 66.5% |
| | 2013/14 | | 10.3% | 37.4% | 52.3% |

### *A.1.3 Seasonal Variation*

An assessment, to determine whether the time of year that the cracking was collected was carried out, to see if this has any effect on the consistency of the data. Data was extracted from the SCANNER database for surveys performed in the summer (between May and September) and winter (between November and February), as shown in Table 10.

**Table 10: Length of data extracted and split by road type**

| | Year | 2010/11 | 2011/12 | 2012/13 | 2013/14 | 2014/15 | 2015/16 |
|---|---|---|---|---|---|---|---|
| Summer (May – Sept) | All data (km) | 71,855 | 63,427 | 71,912 | 91,229 | 89,895 | 82,372 |
| | Rural | 83% | 77% | 84% | 81% | 83% | 80% |
| | Urban | 17% | 23% | 16% | 19% | 17% | 20% |
| | Principal | 39% | 37% | 36% | 38% | 34% | 39% |
| | Non-principal | 61% | 63% | 64% | 62% | 66% | 61% |
| Winter (Nov – Feb) | All data (km) | 19,805 | 16,232 | 25,263 | 8,080 | 12,721 | 10,129 |
| | Rural | 65% | 76% | 70% | 56% | 70% | 82% |
| | Urban | 35% | 24% | 30% | 44% | 30% | 18% |
| | Principal | 38% | 29% | 41% | 52% | 42% | 29% |
| | Non-principal | 62% | 71% | 59% | 48% | 58% | 71% |

As can be seen from Table 10, a lot more data was collected during the summer months than in the winter, with a larger percentage of rural lengths being surveyed in the summer. The distributions of cracking values across the network, for data collected during the summer and also for data collected during the winter, have been plotted and are shown in Figure 24. As can be seen, the cracking data collected during the winter does seem much more variable. This behaviour is not seen in other parameters, such as rutting (Figure 25), where slightly less consistency is seen for winter data but not as much as for cracking. Thus surveying in the winter does not appear to be an issue for SCANNER surveys in general.

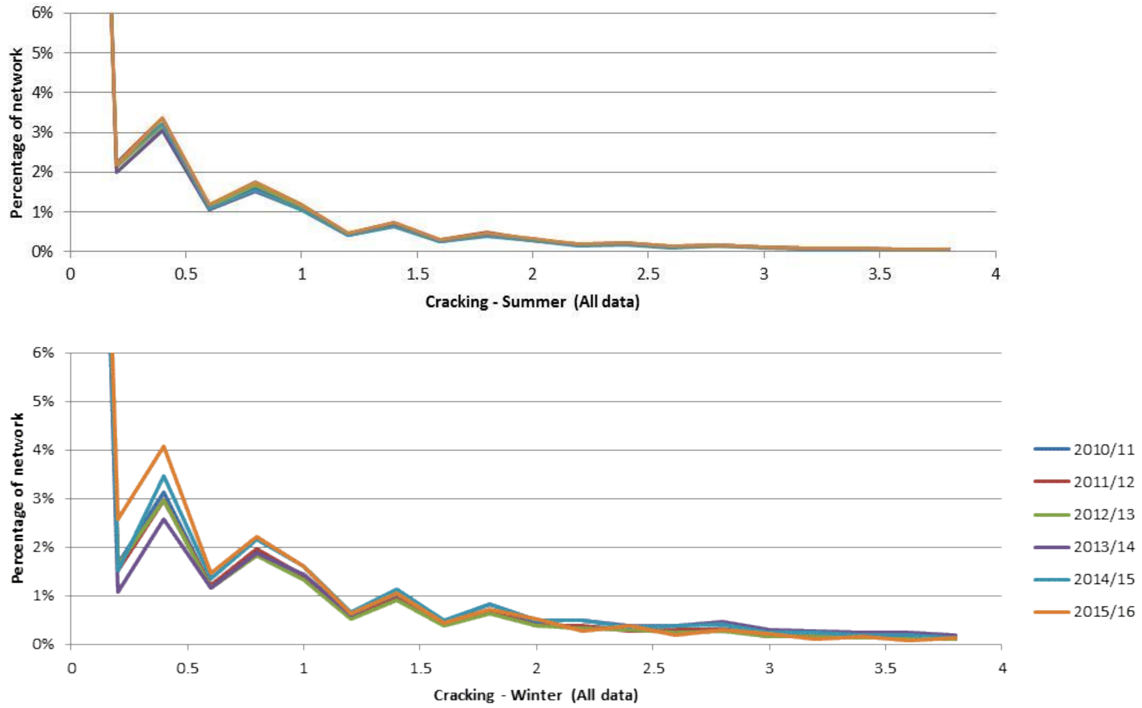



**Figure 24: Distribution of cracking values for surveys performed in the summer (top) and winter (bottom)**

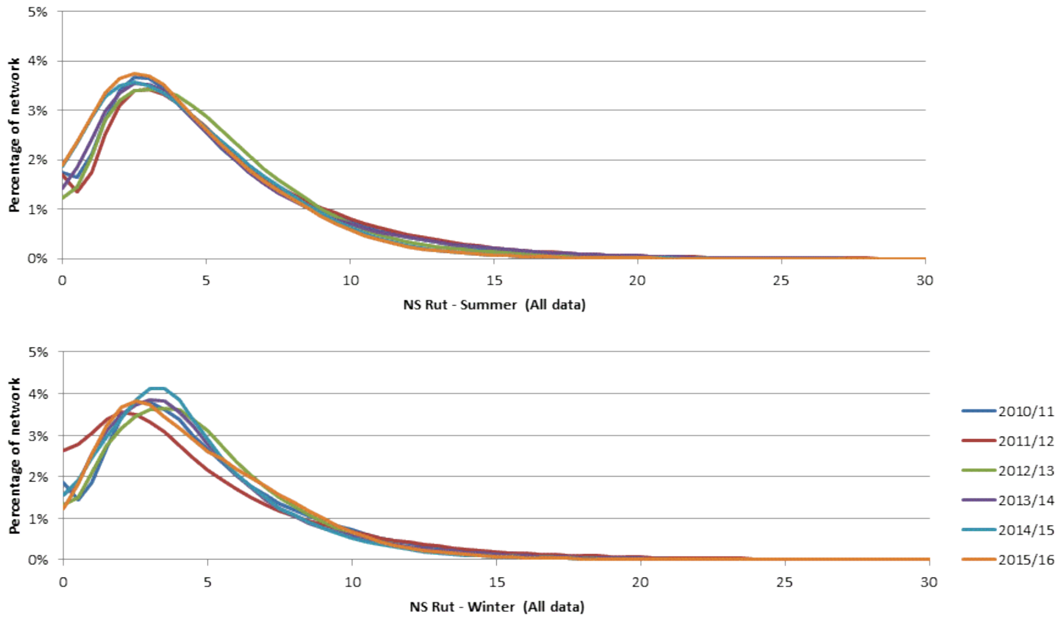



**Figure 25: Distribution of rutting values for surveys performed in the summer and winter**

Since it has been shown that cracking data collected in an urban environment is less consistent than data from a rural environment (Section A.1.2), the larger proportion of urban data contained in the winter surveys may be affecting the consistency. Therefore, data collected in the winter on urban roads has been separated from data collected in the summer and similarly for rural roads. The distributions of this data are shown in Figure 26. As can be seen, the best consistency in the data can be seen in the summer/rural distribution, for which there is very low variability. The greatest variability can be seen in the winter/urban data, with summer/urban and winter/rural being somewhere in between. Since there is variability in the winter/rural data, it can be concluded that the inconsistency seen in the winter data, when all road types are considered, is not solely due to the increased variability due to the urban lengths.

Thus surveying in the winter, or in an urban environment, appears to have a detrimental effect on the consistency of the data.

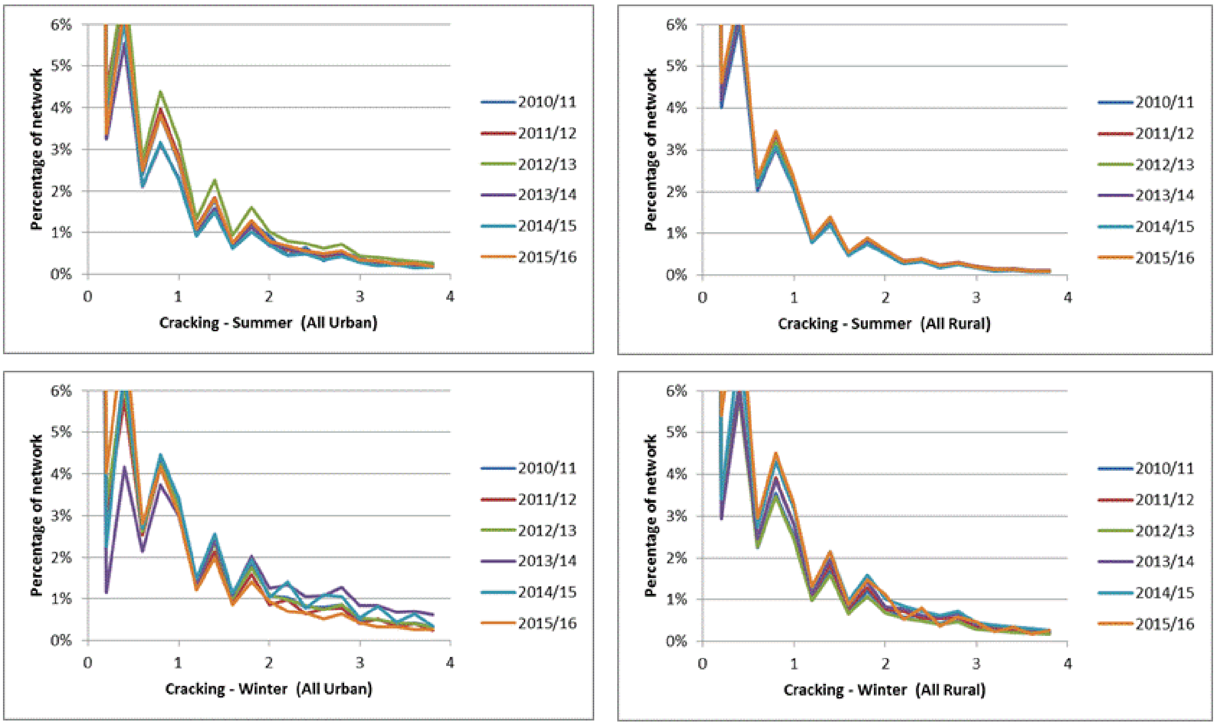



**Figure 26: Distribution of cracking split by survey period and road type**

### A.1.4 Thresholds used for cracking in the RCI calculation

The effect of the RCI thresholds on the consistency of the cracking data was investigated to determine whether the thresholds themselves result in the RCI being over-sensitive to changes in the level of cracking. Data for the 2015/16 survey year was extracted from the SCANNER database and the current thresholds of 0.15 and 2% applied to the data. The percentage of the network contained in each category (Red, Amber, Green) is shown in Table 11. 1.32% of the network has been placed in the Red category for this data. However, this could increase to 1.75% if the inconsistency in the data is taken into account, which is a significant proportional change in the red percentage (the Audit report allows a device to have an average bias between -0.17% and +0.17% and we have assumed a bias of +0.16% in the data to obtain this increase). The Amber category could increase from 61.66% to 64.28%.

**Table 11: Percentage of 2015/16 data lying in each category for cracking and the potential change, due to consistency**

| | Cracking | |
|---|---|---|
| **Category** | **% of network** | **Potential change** |
| **Red** (cracking ≥ 2%) | 1.32% | +0.43% |
| **Amber** (0.15 < cracking < 2%) | 61.66% | +2.62% |
| **Green** (cracking ≤0.15%) | 37.01% | |

To put this into context, the change to the Red and Amber categories that the inconsistency in rutting can cause was also investigated, by simulating changes resulting from rutting inconsistencies on the measured rutting data. Comparing the results in Table 11 and Table 12, it can be seen that the inconsistencies in rutting result in similar changes to the cracking inconsistencies, suggesting the cracking thresholds result in similar absolute change in the red percentage, but the proportional change is greater, as a result of the lower overall percentage red.

**Table 12: Percentage of 2015/16 data lying in each category for rutting and the potential change, due to consistency**

| | Rutting | |
|---|---|---|
| **Category** | **% of network** | **Potential change** |
| **Red** (rutting ≥ 20mm) | 4.96% | +0.69% |
| **Amber** (10 < rutting < 20mm) | 63.61% | +1.89% |
| **Green** (rutting ≤ 10mm) | 31.43% | |

If the upper threshold for a parameter lies on a rapidly changing part of its frequency distribution curve, then very small changes in the parameter value can result in large changes to the number of values exceeding that threshold. Thus this situation results in parameter inconsistency being amplified in the RCI consistency. As can be seen from the distributions shown in Figure 26, the upper threshold for cracking (2%) lies on the part of the distribution where it just starts to flatten out. If this upper threshold were to be decreased, this would result in the threshold lying on a more rapidly changing part of the distribution, which would lead to more inconsistency in the amount of Red lengths reported. Increasing the upper threshold slightly would not change the slope of the distribution significantly and an increase to about 4% would be required for any noticeable difference to be obtained. However, an upper threshold of 4% would result in less than 0.2% of the network being reported as poor, which does not reflect the assessment by engineers or users.

This would suggest that it is not the threshold values used that are causing cracking consistency to have a large effect on the RCI and thus changing the threshold values (within a sensible range) would not help to make the RCI more consistent.

## A.2 Development of tests for Consistency

### *A.2.1 Current fleet*

The data collected during the Accreditation tests for the last 5 years has been collated and the average value on the road routes calculated. These values are plotted in Figure 27.

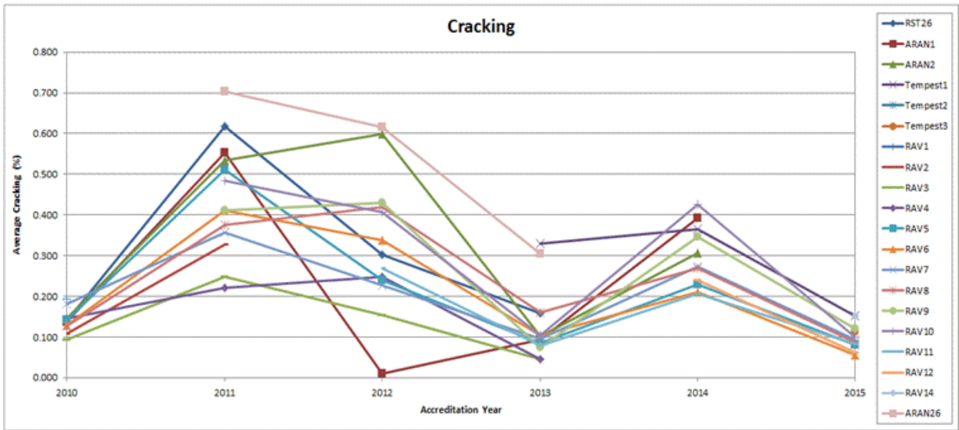



**Figure 27: Average cracking values from road routes included in Accreditation tests**

As can be seen, cracking consistency has improved between the devices in the last year – the range of values is much smaller this year than previous years. However, if individual devices are considered, there is considerable variation from year to year.

The cracking reported in 2015 is ~0.2% lower than that in 2014 and this was reflected in a (smaller) reduction in average value on the network (Section A.1.2).

So, even for the data collected during accreditation (which should be performed in relatively controlled conditions compared to network surveys), there is quite a spread in the values recorded – a difference of about 0.6 in the worst year, which is 4 times the size of the lower threshold used in the RCI calculation.

Also, if the difference between the average value of cracking measured on the SCANNER Road Routes and the machine average value is calculated (Figure 28) there is a difference of 0.22% between the highest average value reported and the lowest, which again is bigger than the lower threshold (0.15%) and >10% of the upper threshold (2%).

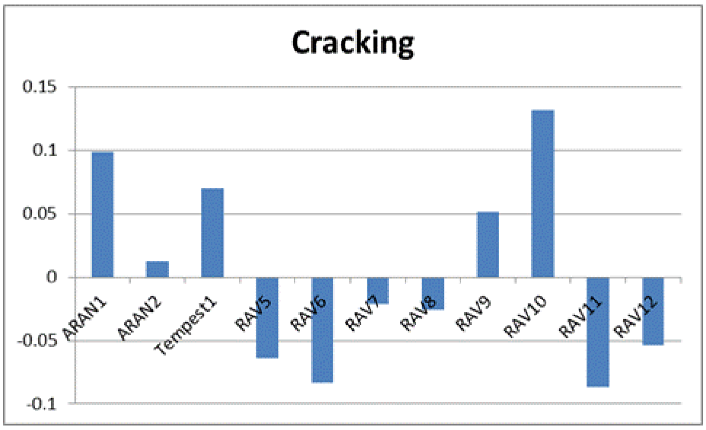



**Figure 28: Difference between the average cracking reported by each device and the machine average value on the road routes**

When nearside rutting (which seems to be the least consistent of the two rutting parameters) is considered, there is a difference of up to 1.5mm, which is significantly less than the 10mm lower threshold applied in the RCI calculation and only 7.5% of the upper threshold.

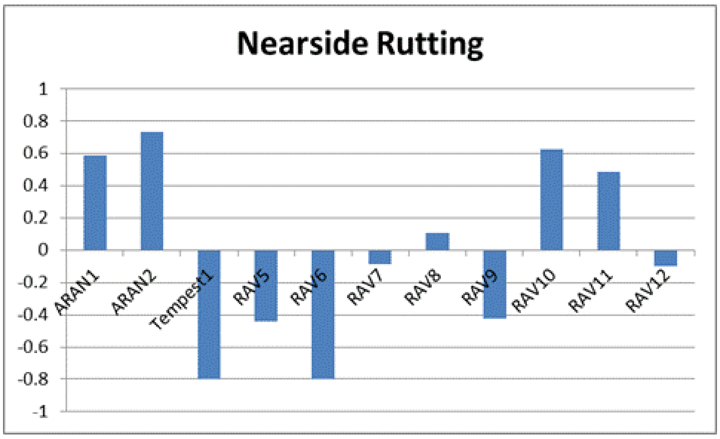



**Figure 29: Difference between the average nearside rutting reported by each device and the machine average value on the road routes**

### *A.2.2 Existing methods*

Several existing methods were identified that had potential to be applied to the repeatability or fleet consistency tests for SCANNER. These were reviewed and assessed for suitability and a summary of each is given in the following sub-sections.

#### *A.2.2.1 Consistency method for SCRIM*

There is currently no method to provide skid resistance reference data and therefore the accreditation test applied to SCRIM devices requires them to all attend an annual assessment where the consistency of the fleet is tested. Any devices being found to be outside of an acceptable range of the rest of the fleet are not accredited and are not allowed to offer surveys.

The method used for SCRIM data is based on calculating the standard deviation between repeat runs from the same device and also standard deviation between the fleet. These are known as the "between run standard deviation" and the "between equipment standard deviation" respectively. Any device determined to not be repeatable or consistent with the rest of the fleet is not awarded the certification required to survey. The standard deviation in each case is calculated as follows:

- **Between Run Standard Deviation**: For several test sections, calculate the mean and the standard deviation of data from repeat runs. If the Between Run Standard Deviation (BRSD) < 3 then the device is considered to be consistent with itself.
- **Between Equipment Standard Deviation:** For several test sections, calculate the mean of the fleet and also the standard deviation of the fleet for each section. If the Between Equipment Standard Deviation (BESD) < 3 then the fleet is considered to be consistent.

#### *A.2.2.2 The Chris Britton method*

The Chris Britton Consultancy (CBC, 2006) devised a method for the DfT that could be used to estimate the overall repeatability, reproducibility and reliability of results obtained from an accredited SCANNER machine.

Each survey machine has a systematic error and random error. The systematic error, known as the bias, can be defined by the deviation from a reference; it is normally predictable as an algebraic offset or a scaling factor. The random error (or precision) is dependent on the length of the sample network and the length of the subsections. The random error decreases with increasing number of subsections. The random error is also dependent on the network characteristics (road class, environment urban or rural, condition for data collection etc.).

CBC proposed that the following equations could be used to calculate these components:
Bias, $\bar{\varepsilon}$= mean result from survey machine – mean result from reference

Error on bias = $\frac{1.96s}{\sqrt{n}}$, and Random error = $\frac{1.96s}{\sqrt{N}}$, where $n$ is the number of subsections in the sample network (e.g. number of 10m lengths), $N$ is the number of subsections in the surveyed network to which the results are being applied and *s* is the standard deviation of the error in the measurement.

Hence
Error = Bias ± confidence on bias ± random error = $\bar{\varepsilon} \pm \frac{1.96s}{\sqrt{n}} \pm \frac{1.96s}{\sqrt{N}}$ (1)

To calculate the errors with respect to a reference, the method calculates the difference between data from the machine and reference data to provide estimates for the bias and the standard deviation. It should be noted that the reference itself has bias and random errors and hence these will falsely be attributed to the machine, but nevertheless this is the best that can be done.

A better estimate for *s* can be obtained by comparing data from repeat surveys of a sample network, compared with the estimate obtained from using a reference, in the sense that errors from the reference are not introduced and the measured variability is attributed to the machine only. To obtain the standard deviation the difference between each pair of measurements is calculated and *s* is then the standard deviation of these differences divided by $\sqrt{n}$ , where *n* is the number of repeat runs available.

To calculate errors between different machines, the same method as for with respect to a reference is used, replacing the reference with data from another device. The errors, calculated using this approach, are then in fact combined errors. If a network has been measured with one machine, then the bound within which the results from the other machine would be expected to lie can be easily estimated.

The factors affecting the calculated errors are the length of the section, and subsection, network characteristics, condition for data collection (weather time of day, light level etc.).

#### A.2.2.3 TRL method

Whilst carrying out an investigation into using the CBC method to calculate the consistency of the BVPI (a network level indicator, based on the percentage of lengths where the RCI≥100), it became apparent that there was an alternative, slightly more straightforward method that could be used for this (Benbow & Wright, 2008). The method assumes a Bernoulli distribution (a special case of a Binomial distribution) for the BVPI. The bias is calculated as the difference of the BVPI for the survey machine and that for the reference machine. The interval of confidence around the bias is calculated as (Agresti and Coull, 1998):

$$CI = \frac{Bias + \frac{1.96^2}{2n} \pm 1.96\sqrt{s^2 + \frac{1.96^2}{4n^2}}}{1 + \frac{1.96^2}{n}} \tag{2}$$

where the standard error of differences, $s = \sqrt{\frac{p_1(1-p_1) + p_2(1-p_2) + 2p_1p_2}{n}}$ (3)

and $p_1$ is the BVPI for the reference machine, $p_2$ the BVPI for the survey machine and n is the number of 10m lengths in the sample network. The method also uses a standard random error calculated as

$$\text{Standard random error} = \sqrt{\frac{p(p-1)}{N}} \tag{4}$$

Where $p$ is the combined proportion of lengths where the RCI exceeds 100 and $N$ is the number of 10m lengths in the surveyed network to which the results are being applied.

#### A.2.2.4 TRACS method

There are two methods included in the TRACS Accreditation tests to assess the repeatability of cracking: The first is identical to that used for SCANNER i.e. comparison of High, Medium and Low levels of cracking, derived from normalised data. The second method considers the differences between the values reported by two runs and again uses normalised data.

Fleet consistency is tested within the TRACS Accreditation by normalising the data and then calculating the distribution of (normalised) cracking values provided by each device:

- The bias of each device from the fleet is obtained by calculating the shift required to obtain the best correlation with the distribution of values from all runs and all devices. The size of this bias/shift is required to be less than 0.05%.

- The cross correlation between the histogram of distributions of intensity data for each device is calculated, following the removal of bias and is required to be at least 0.90.
- Where there are more than two devices, the absolute difference between any of the bias values is calculated. The bias difference must not exceed 0.1%

#### A.2.2.5 Discussion of existing tests

***The SCRIM method:*** This method is based on between run standard deviation and between equipment standard deviation; the method is interesting because a spatial dimension is added to test for the consistency for different locations of the network, which will help improve the efficiency of the accreditation process by identifying length of the network for which the consistency is difficult to achieve. However, a drawback of the method is that a standard deviation is used to quantify the bias between measurements reported by different devices instead of the difference between the mean values recorded.

***The Chris Britton method:*** The method calculates a bias between a reference device and the test device, and a random error between repeat runs of the test device. The random error is calculated as the standard deviation of the differences calculated over the length of the survey. This does not give an ideal estimate as it incorporates the spatial variability in the difference between two repeat runs, and not just the variability between repeat measurements at a single point.

***TRL method:*** The method was devised to calculate the consistency of the BVPI, assuming that the BVPI follows a Bernoulli distribution (a special case of a binomial distribution). It calculates a confidence interval for the difference between (BVPI for the reference machine) and (BVPI for the survey machine). However, it cannot be used to calculate a confidence interval for the difference of proportion, $p_1$-$p_2$.

**TRACS method:** The approaches used to test repeatability within TRACS use normalised data and thus could potentially allow two different runs that are not in the same scale, to be described as repeatable. Whilst this scenario has not been observed so far with any device, it is felt that a method that compares actual and not normalised values, would be of more benefit. The fleet consistency test again uses normalised data, which should work well, where there are likely to be few devices all provided by the same survey contractor (using similar equipment and the same crack detection algorithm) but may not work so well where the spread of data could be much bigger.

Thus it is felt that none of these existing approaches are appropriate to apply to repeatability testing or fleet consistency testing in the SCANNER Accreditation tests.

### A.2.3 Proposed method for repeatability - Confidence Interval and the Coefficient of Variation

#### A.2.3.1 Confidence Interval

Let $X_{ij}^{k}$ be the measurement of the condition of the road obtained for run *i,* subsection *j*, and device *k*. Let m be the total number of runs, n the total number of subsections, and M the total number of devices in a fleet. In general the repeat measurements are not the same and the difference in the measurements is attributed to random errors which calculated as the variance of the repeat measurements:

$$\left(\varepsilon_j^k\right)^2 = \mathrm{var}\left(X_{ij}^k\right) \tag{5}$$

Where $\varepsilon_j^k$ is the between run measurement error for device *k* at subsection *j* and $Var\left(X_{ij}^k\right)$ is the variance of the measurements $X_{ij}^k$ given as:

$$var\left(X_{ij}^k\right) = \frac{1}{m}\sum_{i=0}^{m-1}\left(X_{ij}^k - \overline{X}_j^k\right)^2 \tag{6}$$

Where the term $\bar{X}_j^k$ is the mean of the measurements for device *k* at subsection *j*:

$$\overline{X}_j^k = \frac{1}{m}\sum_{i=0}^{m-1} X_{ij}^k \tag{7}$$

The average of the measurement error for all the n sections is given as:

$$\bar{\varepsilon}_k = \sqrt{\frac{\sum_{j=0}^{n-1}\left(\varepsilon_j^k\right)^2}{n}} \tag{8}$$

Equation 8 gives the typical measurement error that quantifies the spread of the repeat measurement around the true measure for device k. It is assumed that the measurement error $\bar{\varepsilon}_k$ is the contribution of equal independent individual errors from the m repeat runs, and thus the individual error e is defined as:

$$e_k = \frac{\bar{\varepsilon}_k}{\sqrt{m}} \tag{9}$$

For the case where the number of repeat runs is m=2, Equation 5 becomes

$$e_k = \frac{\bar{\varepsilon}_k}{\sqrt{2}} \tag{10}$$

and the measurement error $\bar{\varepsilon}_k$ is given as

$$\bar{\varepsilon}_k = \sqrt{\frac{\sum_{j=0}^{n-1}\left(X_{2j}^k - X_{1j}^k\right)^2}{2n}} \tag{11}$$

It is assumed that the distribution of all the measurement errors produced during the life of the device follow a normal distribution; hence it is possible to construct a 95% confidence interval around the best estimate of the population average given as:

$$ci_k = \pm 1.96 e_k \tag{12}$$

If there is a bias in the repeat data, this is known as the internal bias (to distinguish from the bias between different devices).  Whilst the internal bias is usually equal to zero, in the case that it is not, let $\beta_k$ be this bias. The confidence interval is now given as:

$$ci_k = \pm\left(\beta_k + 1.96 e_k\right) \tag{13}$$

Figure 30 shows a plot of the standard deviation $\epsilon_j^k$ of cracking data against its mean $\bar{X}_j^k$, calculated using accreditation data from the SCANNER road routes. Visual assessment of this plot would suggest that the standard deviation is related to the magnitude of the cracking data and this can be confirmed using the Kendal tau rank correlation test (Kendal $\tau = 0.83$ and correlation coefficient $\rho = 0.70$).

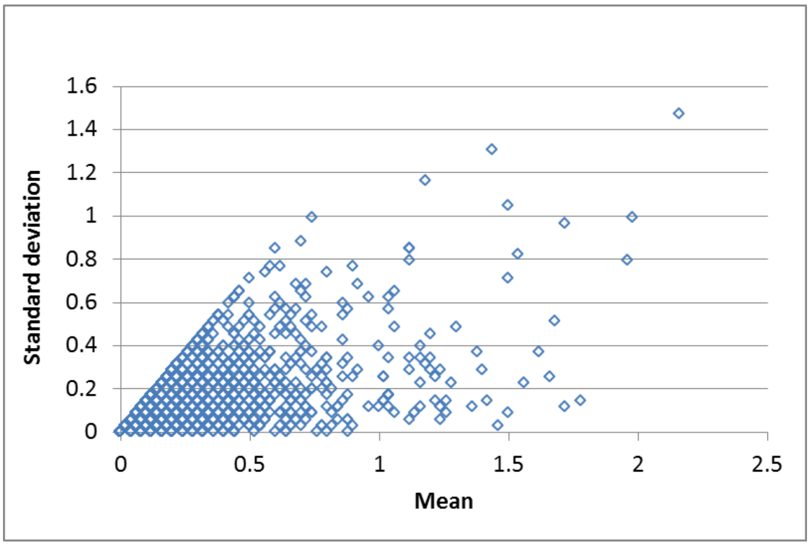



**Figure 30: Plot of the standard deviation of the data against the mean of the data**

A confidence interval of the form given by Equation 13 assumes that mean and standard deviation are not correlated and therefore it is not appropriate to apply this directly to the cracking data. A fix for this is to log-transform the cracking data (using a base 10 logarithm), and Figure 31 shows the standard deviation against the mean when log-transformed data is used. Application of the Kendal tau rank correlation test to this data demonstrates that the standard deviation and the mean are not correlated (Kendal $\tau$=-0.13 and correlation coefficient $\rho$=-0.20).

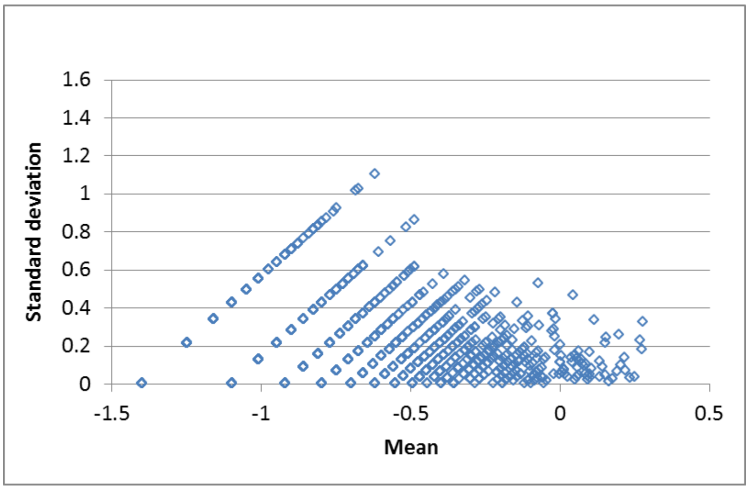



**Figure 31: Plot of the standard deviation of the log-transformed data against the mean of the log-transformed data**

Thus Equation 13 can be applied to calculate a confidence interval for cracking in the log scale. Since it is more practical to calculate a confidence interval in the original scale, this can be calculated as:

$$ci_k = \overline{X_k} \times 10^{\pm(1.96e_k)}, \quad (14)$$

Where $\overline{X_k}$ is the average value of cracking reported by device k and it has been assumed that the bias between repeat runs from the same device is zero i.e. $\beta_k$=0.

Whilst the confidence interval, given in Equation 14, could be used, the width of the confidence interval has more meaning as a parameter for repeatability, in the sense that the narrower the width the more confidence we have in the data, and the wider it is the less confidence we have in the data. The width of the confidence interval is calculated as:

$$CI_k = \overline{X}_k \frac{1-(ci_k)^2}{ci_k} \quad (15)$$

### *A.2.3.2 Coefficient of Variation*

The coefficient of variation is formally defined as the ratio of the standard deviation, σ to the mean, μ. An unbiased estimate of the coefficient of variation is given for a normal distribution as:

$$CV = \left(1 + \frac{1}{4n}\right)^2 \frac{\sigma}{\mu} \quad (16)$$

Since the cracking data is log-transformed in the calculation of Confidence Interval (Section A.2.3.1), the Coefficient of Variation (CV) for a lognormal distribution has been used, which expresses the variability in the data as a number between 0 and 1. This is defined as:

$$CV = \sqrt{e^{(\sigma \ln(10))^2} - 1} \quad (17)$$

An advantage of using CV to assess repeatability is that it is not affected by scale and hence is suitable to use for assessing the consistency of data across all the parameters, which could enable e.g. the repeatability of rutting to be directly compared with that for cracking.

However, because CV is a relative measure, it can make the repeatability look poor for lengths where the magnitude of cracking is small (a regular occurrence, since ~60% of the network is reported to have no cracking at all). Small absolute differences in cracking are not considered to be of concern, even when the values of cracking are small themselves. Therefore, a further step is added, after the calculation of CV (using Equation 17): *If the value of CV exceeds 0.1 but the mean value for all runs is less than 0.1 and the standard deviation is less than 0.03, then set the value of CV=0.05*.

### *A.2.3.3 Length of subsection over which to calculate CI and CV*

The SCANNER parameters are reported over 10m lengths and the consistency parameters are usually calculated for this interval too. However, during Accreditation, the parameters are averaged over 50m lengths, to reduce the effect of location referencing errors on the data. The length over which to calculate CI and CV has therefore been considered, to determine whether 10m, 50m or a longer aggregation length would be most appropriate to use.

Figure 32 and Figure 33 show that the site average for the CI and CV (respectively) decrease with increasing subsection length, but stabilise in value after an aggregation length of 500m. The longer the aggregation length used, the shorter the processing time to calculate the values and the easier the data is to handle, when performing further investigation. Thus an aggregation length of 500m has been used herein.

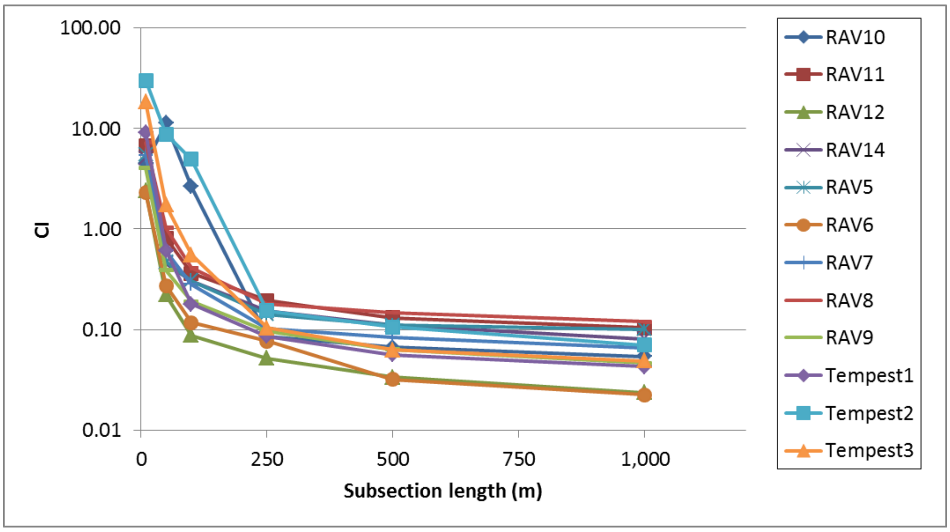



**Figure 32: Effect of subsection length on the value of CI**

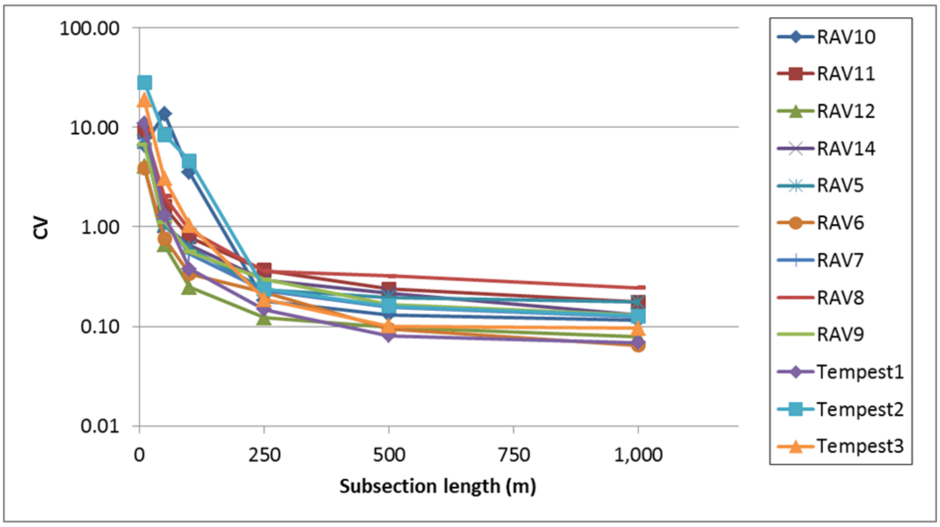



**Figure 33: Effect of subsection length on the value of CV**

### *A.2.4 Proposed methods for fleet consistency*

#### *A.2.4.1 Centre of Gravity Method*

The centre of gravity of a frequency distribution could be used to provide a representative value for the fleet and is defined in Equation 18

$$CG = \frac{\sum_{i=1}^{n} x_i p_i}{\sum_{i=1}^{n} p_i} \quad \textbf{(18)}$$

Where $x_i$ is height of the distribution graph at distance $p_i$.

For normally distributed data, the CofG is equivalent to the mean value of the data. Similarly, for any distribution, as the size of ranges used for the distribution decreases, the CofG tends towards the mean.

#### *A.2.4.2 Percentile Range Method*

A representative value could also be obtained by taking the mean of values lying between an upper bound Sup and a lower bound Inf, defined as:

- $Sup = 50th\ percentile + 1.5\ (\ 75th\ percentile - 25th\ percentile)$
- $Inf\ = 50th\ percentile - 1.5\ (\ 75th\ percentile - 25th\ percentile)$

If $V_1$,…$V_n$ are the values provided by the fleet, in ascending order of value, then the representative value = mean($V_p$,…,$V_q$) where $V_p \geq Inf$ and $V_q \leq Sup$.

#### *A.2.4.3 Cluster Method*

The cluster method determines a representative value by determining which devices report values that are close together (i.e. which devices are clustered) and then calculates a mean of these clustered values.

If $V_1$,…$V_n$ are the values provided by the fleet, in ascending order of value, then

- Calculate $V_2$-$V_1$.
  - If $V_2$-$V_1$<0.1, then these can be considered to be in Cluster 1
  - Otherwise $V_1$ is a single point
- Calculate $V_3$-$V_2$
  - If this is <0.1, then $V_3$ is in the same cluster as $V_2$
  - Otherwise $V_3$ is in a separate cluster to $V_2$.
- Continue this process until you run out of data points.
- Determine whether any of the clusters include more than 50% of the points.
  - If not, join together any clusters that have a gap of <0.2
  - This is the "representative cluster".
- Calculate the mean of the "representative cluster".

### *A.2.5 Assessing the methods to calculate a representative value*

To assess how well each of the representative value methods works, the procedure set out in Section 2.5.1, and the thresholds determined in Section 2.5.3, have been applied to several datasets:

- Data from the devices Accredited between April and December 2016
- Data from the devices Accredited between April and December 2016 plus one simulated outlier with values larger than the fleet, MU1 = 5* RAV6 values
- Data from the devices Accredited between April and December 2016 plus three simulated outliers with values larger than the fleet, MU1 = 5* RAV6 values, MU2 = 4*RAV6 values, MU3 = 3*RAV6 values
- Data from the devices Accredited between April and December 2016 plus one simulated outlier with values smaller than the fleet, MU4 = 0.25* RAV6 values

### A.2.5.1 Centre of Gravity results

The results shown in Table 13 show that using the centre of gravity to obtain a representative value for the fleet works well when the fleet is relatively clustered and there are no outliers. However, the addition of one or more outliers can change the results significantly e.g. RAV8 passes until any of the outliers are added (MU1,…,MU4). It is interesting to note that none of the devices pass the test when three vehicles delivering large values (MU1 to MU3). It's likely that this is because the method calculates a representative value that lies in the middle of the two halves of the fleet.

The method does, however, fail all of the added outlying devices.

**Table 13: Results of fleet consistency tests using centre of gravity to obtain representative value**

| Vehicle | Data | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | April-Dec 2016 Accreditation devices* | | | 2016 + MU1 | | | 2016 + MU1,MU2,MU3 | | | 2016 + MU4 | | |
| | 1 | 2 | Pass? | 1 | 2 | Pass? | 1 | 2 | Pass? | 1 | 2 | Pass? |
| RAV6 | 0.040 | 46.5% | N | 0.044 | 45.5% | N | 0.040 | 41.4% | N | 0.037 | 52.5% | N |
| RAV7 | 0.018 | 90.9% | Y | 0.025 | 81.8% | Y | 0.046 | 46.5% | N | 0.026 | 82.8% | Y |
| RAV8 | 0.024 | 84.8% | Y | 0.056 | 50.5% | N | 0.064 | 36.4% | N | 0.058 | 50.5% | N |
| RAV9 | 0.041 | 59.6% | N | 0.047 | 59.6% | N | 0.056 | 30.3% | N | 0.048 | 56.6% | N |
| RAV10 | 0.057 | 51.5% | N | 0.041 | 53.5% | N | 0.060 | 35.4% | N | 0.038 | 46.5% | N |
| RAV11 | 0.061 | 56.6% | N | 0.020 | 80.8% | Y | 0.043 | 45.5% | N | 0.016 | 85.9% | Y |
| RAV12 | 0.045 | 57.6% | N | 0.024 | 79.8% | Y | 0.046 | 39.4% | N | 0.021 | 82.8% | Y |
| RAV14 | 0.024 | 77.8% | Y | 0.041 | 58.6% | N | 0.060 | 22.2% | N | 0.035 | 69.7% | Y |
| RAV15 | 0.037 | 57.6% | N | 0.029 | 73.7% | Y | 0.054 | 34.3% | N | 0.017 | 86.9% | Y |
| MU1 | - | - | - | 0.249 | 10.1% | N | 0.223 | 9.1% | N | - | - | - |
| MU2 | - | - | - | - | - | - | 0.161 | 15.2% | N | - | - | - |
| MU3 | - | - | - | - | - | - | 0.102 | 24.2% | N | - | - | - |
| MU4 | - | - | - | - | - | - | - | - | - | 0.043 | 41.4% | N |

1 = Average Bias for site

2 = Percentage of 500m lengths for which the absolute bias <0.036

* Note that Yotta devices have not been included, as, at the time of analysis, the 2016/17 Accreditation data was not available

### A.2.5.2 Percentile range method results

The results are similar to the centre of gravity method, when a percentile range is used to obtain the representative value (Table 14), with inconsistencies as to whether devices pass or fail the test e.g. RAV6, RAV8, RAV10, RAV11, RAV12. Also, whilst the large outliers (MU1 to MU3) fail the test, the small valued device (MU4) passes.

**Table 14: Results of fleet consistency tests using percentile range to obtain representative value**

| Vehicle | Data | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | April-Dec 2016 Accreditation devices | | | 2016 + MU1 | | | 2016 + MU1,MU2,MU3 | | | 2016 + MU4 | | |
| | 1 | 2 | Pass? | 1 | 2 | Pass? | 1 | 2 | Pass? | 1 | 2 | Pass? |
| RAV6 | 0.039 | 65.7% | Y | 0.039 | 63.6% | N | 0.041 | 56.6% | N | 0.037 | 69.7% | Y |
| RAV7 | 0.016 | 91.9% | Y | 0.026 | 84.8% | Y | 0.025 | 85.9% | Y | 0.031 | 82.8% | Y |
| RAV8 | 0.029 | 85.9% | Y | 0.060 | 56.6% | N | 0.053 | 55.6% | N | 0.064 | 54.5% | N |
| RAV9 | 0.047 | 59.6% | N | 0.050 | 59.6% | N | 0.045 | 61.6% | N | 0.054 | 57.6% | N |
| RAV10 | 0.064 | 50.5% | N | 0.035 | 56.6% | N | 0.039 | 54.5% | N | 0.033 | 77.8% | Y |
| RAV11 | 0.069 | 54.5% | N | 0.013 | 93.9% | Y | 0.017 | 85.9% | Y | 0.014 | 91.9% | Y |
| RAV12 | 0.051 | 59.6% | N | 0.016 | 88.9% | Y | 0.021 | 76.8% | Y | 0.016 | 87.9% | Y |
| RAV14 | 0.016 | 84.8% | Y | 0.031 | 81.8% | Y | 0.037 | 70.7% | Y | 0.027 | 90.9% | Y |
| RAV15 | 0.030 | 68.7% | Y | 0.015 | 91.9% | Y | 0.022 | 80.8% | Y | 0.013 | 93.9% | Y |
| MU1 | - | - | - | 0.265 | 9.1% | N | 0.255 | 8.1% | N | - | - | - |
| MU2 | - | - | - | - | - | - | 0.192 | 11.1% | N | - | - | - |
| MU3 | - | - | - | - | - | - | 0.131 | 22.2% | N | - | - | - |
| MU4 | - | - | - | - | - | - | - | - | - | 0.035 | 66.7% | Y |

1 = Average Bias for site

2 = Percentage of 500m lengths for which the bias <0.1

### *A.2.5.3 Clustering method results*

When the clustering method is used, the results obtained are much more consistent, with fewer devices jumping between passing and failing: Only 4 devices change, compared to 5 using the percentile range method and 6 using the centre of gravity method. Also, all outliers fail the test (Table 15).

The behaviour of the method for RAV15, when considering only data from the actual fleet, is slightly odd, in that the device passes the test when considering the average bias for the site (it is less than 0.036) but the bias is ≤0.036 for only 56.6% of the 500m lengths. If one considers RAV11, which also has 56.6% of lengths with a bias ≤0.036 but a much higher average values, it can be seen that this device has a similar number of lengths with a bias in the range 0-0.012 but many more that are >0.06 than RAV15 (Figure 34). This is a phenomenon that could be seen in all of the methods considered herein.

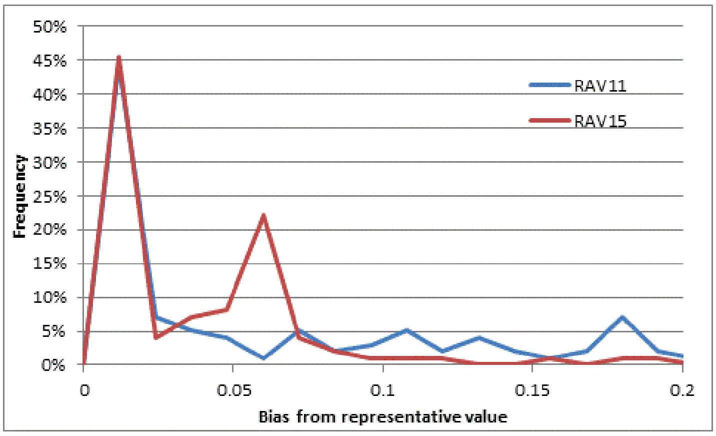



**Figure 34: Frequency distribution of biases for 500 lengths for RAV11 and RAV15**

**Table 15: Results of fleet consistency tests using clustering method to obtain representative value**

| Vehicle | Data | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | April-Dec 2016 Accreditation devices | | | 2016 + MU1 | | | 2016 + MU1,MU2,MU3 | | | 2016 + MU4 | | |
| | 1 | 2 | Pass? | 1 | 2 | Pass? | 1 | 2 | Pass? | 1 | 2 | Pass? |
| RAV6 | 0.039 | 56.6% | N | 0.041 | 54.5% | N | 0.044 | 44.4% | N | 0.037 | 63.6% | N |
| RAV7 | 0.016 | 90.9% | Y | 0.025 | 80.8% | Y | 0.032 | 75.8% | Y | 0.027 | 85.9% | Y |
| RAV8 | 0.024 | 84.8% | Y | 0.056 | 52.5% | N | 0.057 | 54.5% | N | 0.058 | 54.5% | N |
| RAV9 | 0.042 | 59.6% | N | 0.044 | 62.6% | N | 0.048 | 53.5% | N | 0.049 | 57.6% | N |
| RAV10 | 0.059 | 50.5% | N | 0.038 | 51.5% | N | 0.043 | 54.5% | N | 0.038 | 50.5% | N |
| RAV11 | 0.061 | 56.6% | N | 0.018 | 86.9% | Y | 0.027 | 69.7% | Y | 0.014 | 90.9% | Y |
| RAV12 | 0.046 | 58.6% | N | 0.023 | 79.8% | Y | 0.031 | 72.7% | Y | 0.020 | 87.9% | Y |
| RAV14 | 0.022 | 81.8% | Y | 0.041 | 60.6% | N | 0.049 | 46.5% | N | 0.033 | 82.8% | Y |
| RAV15 | 0.035 | 56.6% | Y | 0.024 | 78.8% | Y | 0.036 | 55.6% | Y | 0.014 | 92.9% | Y |
| MU1 | - | - | - | 0.254 | 10.1% | N | 0.241 | 10.1% | N | | | - |
| MU2 | - | - | - | | | - | 0.178 | 17.2% | N | | | - |
| MU3 | - | - | - | | | - | 0.118 | 31.3% | N | | | - |
| MU4 | - | - | - | | | - | - | - | - | 0.041 | 42.4% | N |

1 = Average Bias for site

2 = Percentage of 500m lengths for which the bias <0.036

Since it has been shown to be more robust and consistent, it is felt that the cluster method is the most appropriate method to apply for fleet consistency testing.

# Appendix B Consistency of SCANNER Rutting data (Task 1)

## B.1 Understanding the consistency

### *B.1.1 Year on year consistency*

The Accreditation data from tests carried out in 2014 and 2015 have been used to look at consistency of rutting data by determining how consistently devices report lengths in the same RCI category i.e. are lengths, where rutting is Green in 2014, also reported to be Green in the 2015 data. The rutting did not exceed the Amber/Red threshold (20mm) anywhere on the Accreditation sites and thus it has only been possible to investigate the consistency of reporting Green or Amber with this data. Since the repeatability requirement for rutting is that 95% of lengths are within 3mm, we have considered the ranges 0-7mm, 7-10mm, 10-13mm and 13-20mm. Two representative examples of the results of this are given in Table 16.

**Table 16: Repeatability of nearside rutting data between years for Accreditation data**

| **Tempest 1** | | **2015** | |
|---|---|---|---|
| **2014** | **Total** | **x≤10mm** | **10<x<20mm** |
| **x≤7mm** | 92.52% | **99.50%** | 0.50% |
| **7<x≤10mm** | 6.35% | **90.00%** | 10.00% |
| **10<x≤13mm** | 0.95% | 51.85% | **48.15%** |
| **13mm <x** | 0.18% | 0.00% | **100.00%** |

| **RAV8** | | **2015** | |
|---|---|---|---|
| **2014** | **Total** | **x≤10mm** | **10<x<20mm** |
| **x≤7mm** | 84.05 | **99.54%** | 0.46% |
| **7<x≤10mm** | 12.67% | **90.81%** | 9.19% |
| **10<x≤13mm** | 2.82% | 43.75% | **56.25%** |
| **13mm <x** | 0.46% | 23.09% | **76.92%** |

If rut depths were perfectly consistent, it would be expected that all values in bold in Table 16 would be 100% but, as can be seen, this is not the case and the performance for values in the range 10 to 13mm is particularly poor: Almost 52% of the lengths reported with values between 10-13mm in 2014 were reported as ≤10mm in 2015 for Tempest 1.

Whilst this does suggest that the data is inconsistent to an extent that it will affect the RCI calculation, less than 1% of the network as a whole is affected by lengths contributing to the RCI one year but not the next, or vice versa. So, it is not a large problem on the routes surveyed during Accreditation.

### *B.1.2 Fleet consistency*

Figure 35 shows the average rutting value reported by each device for the last 6 years on SCANNER Road Route 2. It can be seen that the fleet is more consistent in 2015 than it was in previous years, with the nearside rut depths being slightly less consistent than the offside. The largest range seen between the device reporting lowest levels of rutting and that reporting the highest level of rutting in one year is about 3.5mm, whereas the range for 2015 is 1mm. So, the range of values is between 10 and 35% of the lower RCI threshold (10mm) for rutting. This shows that the fleet consistency is good for rutting, particularly when compared to cracking, where the maximum range was 0.6% - four times the lower RCI threshold and 0.15% for 2015 equal to the lower RCI threshold. Thus this would suggest that the fleet is fairly consistent for rutting.

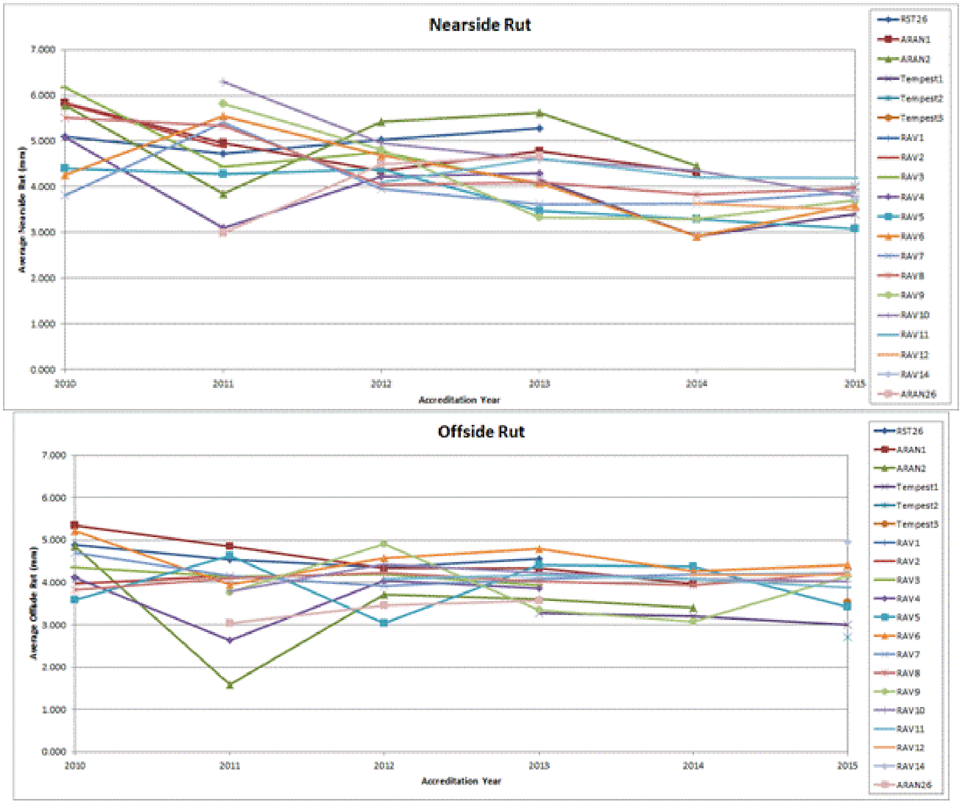



**Figure 35: Fleet consistency from Accreditation data**

Whilst overall fleet consistency is ok, there is however a noticeable difference between the two current contractors, with Yotta reporting an average rut depth of 1.7mm less than WDM (Figure 4). Since the contractors implement their own rut depth algorithm, this difference could be due to a difference in the measurement of transverse profile between the contractors or a difference in the algorithms implemented.

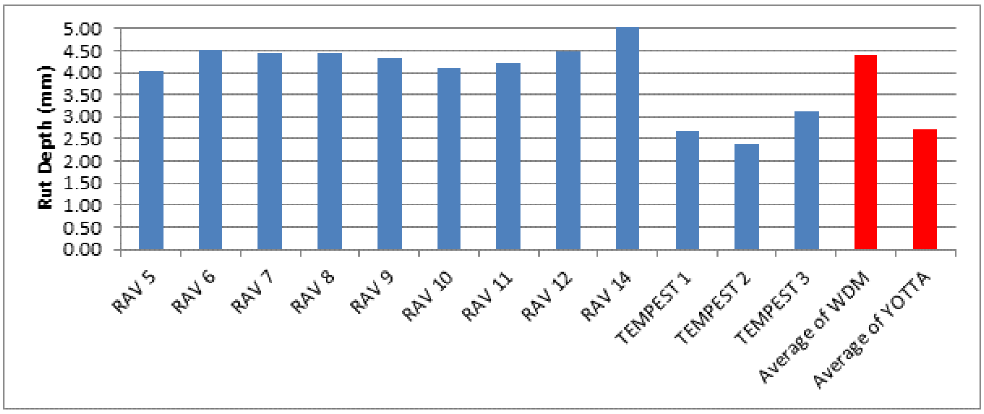



**Figure 36: Average offside rut depths from each device (2014 and 2015 data)**

To investigate this, the transverse profiles from each device were processed through TRL's bespoke software and analysed (Section B.1). It was found that:

- When rut depths are calculated from the raw transverse profile, using the Highways England's TRACS algorithm, the difference between the two contractors can still be seen. This suggests that it is not the contractor's algorithms causing the main difference.

- The differences in rut depth, on lengths where WDM report higher rut depth values than Yotta, are mainly caused by a different driving line taken by the two contractors: the Yotta devices tend to drive further to the left when compared to WDM devices. Thus both contractors are measuring the transverse profile similarly but driving line is causing large differences in these cases.

To investigate the difference seen between the average rutting values reported by the two contractors, the transverse profiles from each device were processed through TRL's bespoke software, which can be used to apply the Highways England algorithm for rutting to the data. As can be seen, whilst this brings the values provided by the contractors slightly closer together – 1.5mm, it does not make a significant difference. This suggests that the difference is mainly due to the transverse profiles being measured by the two contractors, not the rut depth calculation used.

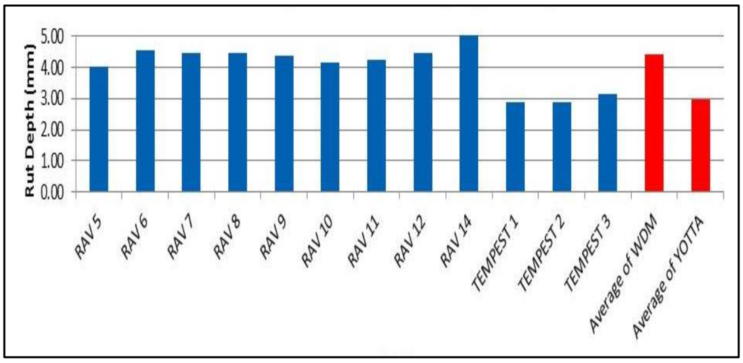



**Figure 37: Average offside rut depths from the fleet, processed using the TRACS rutting algorithm**

The average value could be affected by spikes in the data thus the offside rutting data, reported over 10m lengths, was plotted on a graph for the whole of SCANNER Road Route 2. The graphs, shown in Figure 38, are representative of the whole site and, as can be seen, none of the data contains spikes but the Yotta data is, in general, lower than the WDM data.

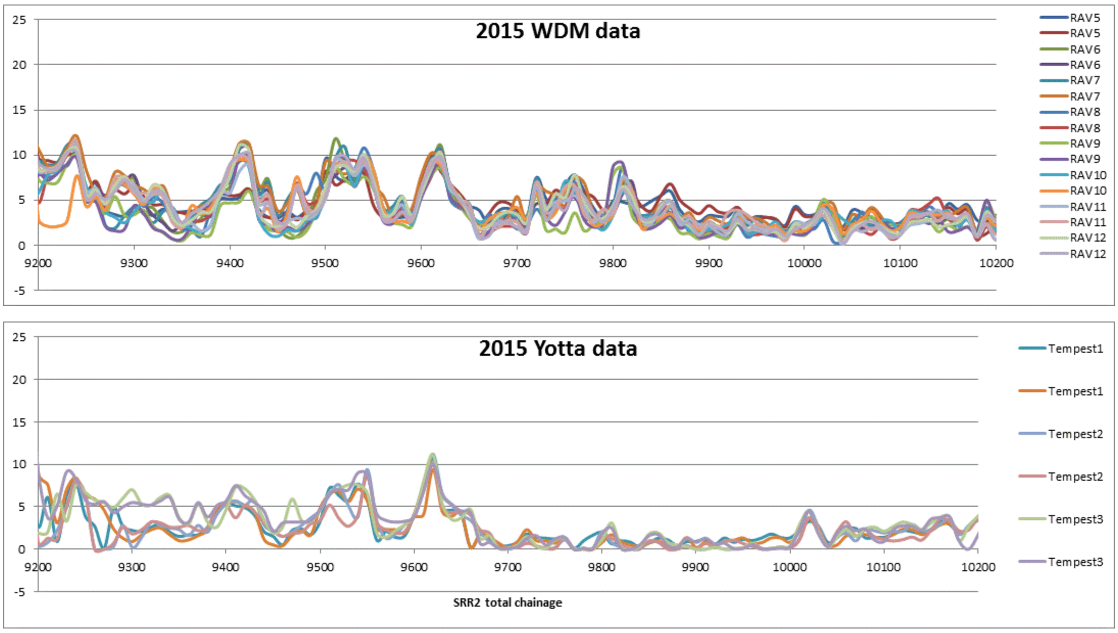



**Figure 38: Rutting data from 2015 surveys of SRR2 (run 1 and run 2 for each device)**

If the two fleets are different for rutting values <10mm but similar for those >10mm, then it could be considered that this is not too much of a problem, as it will not affect the RCI calculation. Therefore, the average value for each device, where the reference has reported a rut value of ≤10mm and the average where the reference >10mm were calculated, for each device. The results of this are presented in Figure 39 and it can be seen that for lengths where the reference reports ≤10mm Yotta devices report ~1.5mm less than WDM but for lengths where the reference reports >10mm (i.e. those that would contribute to the RCI score), the difference is almost 3mm. Thus the issue is actually worse for the lengths where it matters.

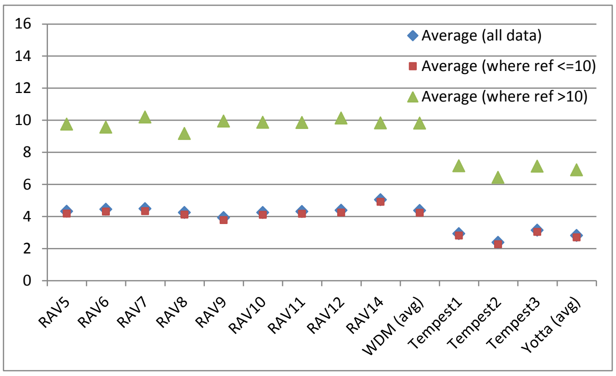



**Figure 39: Average values for SCANNER fleet where the reference is below or above the Green/Amber RCI threshold (10mm)**

Whether the two fleets report Green, Amber and Red lengths consistently has been investigated by comparing the categorisation of each 10m length by the average of the WDM fleet with that of the Yotta fleet. As can be seen from Table 17, over 97% of the network was categorised as Green (≤10mm) by both fleets, whilst none of the network was categorised as Red (>20mm).

**Table 17: Comparison of the average categorisation of individual 10m lengths on the network by the WDM and Yotta fleets**

| | | Yotta | | |
|---|---|---|---|---|
| | | **G** | **A** | **R** |
| **WDM** | **G** | 97.56% | 0.18% | 0.00% |
| | **A** | 1.84% | 0.42% | 0.00% |
| | **R** | 0.00% | 0.00% | 0.00% |

The lengths where WDM report higher rut depth values than Yotta have been investigated further by inspection of the transverse profiles from these lengths. For all the lengths inspected, the differences in rut depth are mainly caused by a different driving line taken by the two contractors: the Yotta devices tend to drive further to the left when compared to WDM devices – for example the transverse profiles shown in Figure 40, where Tempest 1 data has been shifted by 700mm in order to roughly align with RAV14 data. Thus both contractors are measuring the transverse profile similarly but driving line is causing large differences in these cases.

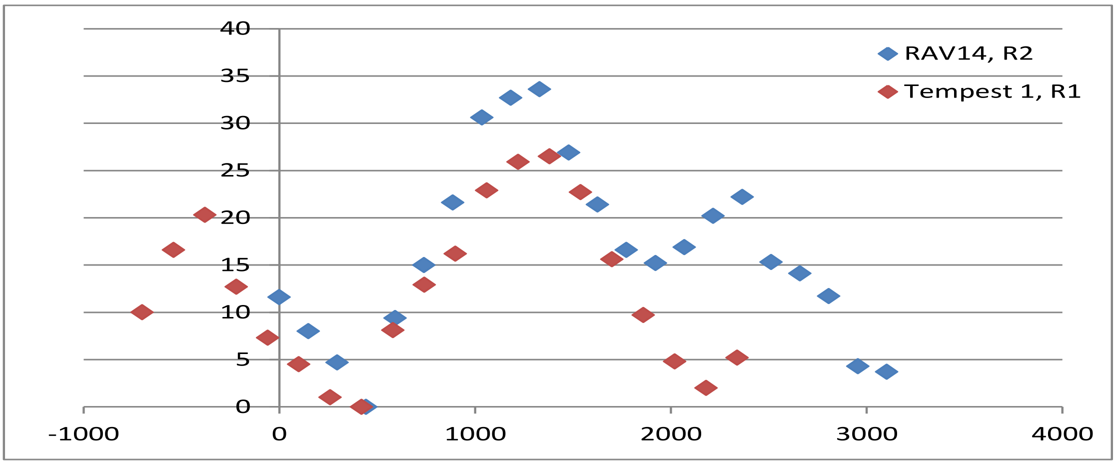



**Figure 40: Transverse profiles from Section 33 of SRR2, chainage 2759.5m**

## B.2 Task 1: Approaches to Improve Rutting Consistency

### *B.2.1 Cleaned rutting*

Cleaned rutting was added to SCANNER along with several other enhanced parameters in 2007. Instead of the contractors implementing their own algorithm, cleaned rutting is calculated using a centrally defined algorithm that attempts to identify the edge of the road in the data and exclude any points made outside of this edge to be excluded from the rut depth calculation i.e. the calculation is based on a “cleaned” transverse profile.

Whether the use of cleaned rutting would improve consistency has been investigated by assessing the effect on a few LHAs when replacing standard rutting with cleaned rutting. The results of this, using data from Norfolk are shown in Table 18. As can be seen, using cleaned rutting results in larger biases between the two years and no improvements in the consistency of RCI categorisation.

There is also a poorer correlation between the distributions (lower distribution correlation numbers), consistent with a visual assessment of the frequency distribution graphs, where a much more inconsistent shape can be seen (Figure 41).

**Table 18: Effect on Norfolk Audit report of using cleaned rutting**

| Rut calculation | Parameter | Year | Mean | 99th percentile | Distn. Corr. | Bias from Prev. | Red | Amber | Green |
|---|---|---|---|---|---|---|---|---|---|
| Standard rutting | N/S rut | 2015/16 | 5.53 | 14.1 | 0.96 | -0.28 | 0.10% | 9.00% | 91.00% |
| | | 2013/14 | 5.81 | 20.4 | | | 1.10 | 13.80% | 85.00% |
| | O/S rut | 2015/16 | 5.59 | 13.7 | 0.98 | 0.24 | 0.00 | 9.40% | 90.60% |
| | | 2013/14 | 5.35 | 16.3 | | | 0.30 | 9.60% | 90.20% |
| Cleaned rutting | N/S rut | 2015/16 | 5.27 | 18.6 | 0.84 | 0.74 | 0.70 | 9.10% | 90.20% |
| | | 2013/14 | 4.53 | 23.5 | | | 1.70 | 8.30% | 90.00% |
| | O/S rut | 2015/16 | 6.1 | 18.6 | 0.94 | 1.29 | 0.70 | 13.80% | 85.50% |
| | | 2013/14 | 4.82 | 16.7 | | | 0.40 | 7.90% | 91.70% |

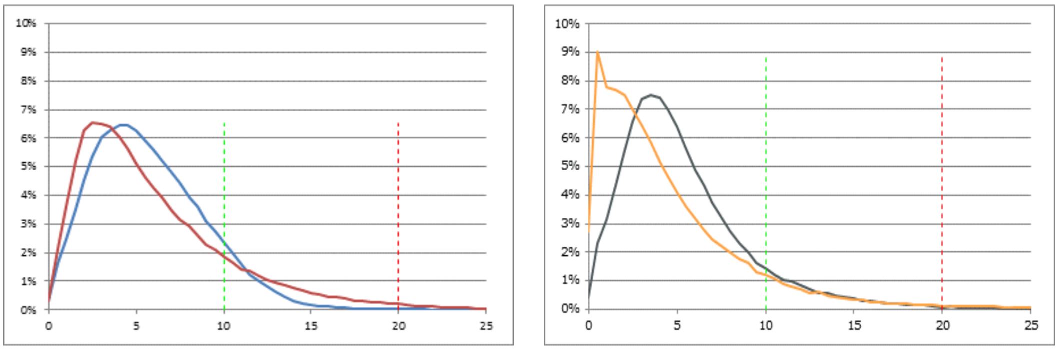



**Figure 41: Nearside rutting frequency distributions: Standard algorithm (left), Cleaned algorithm (right). Blue and grey lines are for 2015/16 data, red and orange 2013/14 data**

This reduction in performance when using cleaned rutting was not expected as this parameter was developed to try and improve the measurement of rutting. So an investigation into what might be causing this was carried out.

The first thing investigated was the identification of the road edge by the rutting algorithms, as it had been noted that a lot of the issues seen with the cleaned rutting algorithm coincided with poor edge detection.

In order to obtain reference data for the position of the edge, the downward and forward facing video has been manually analysed to identify the edge of the lane. Figure 42 shows about 100m road that has been manually analysed and where this has identified the nearside edge.

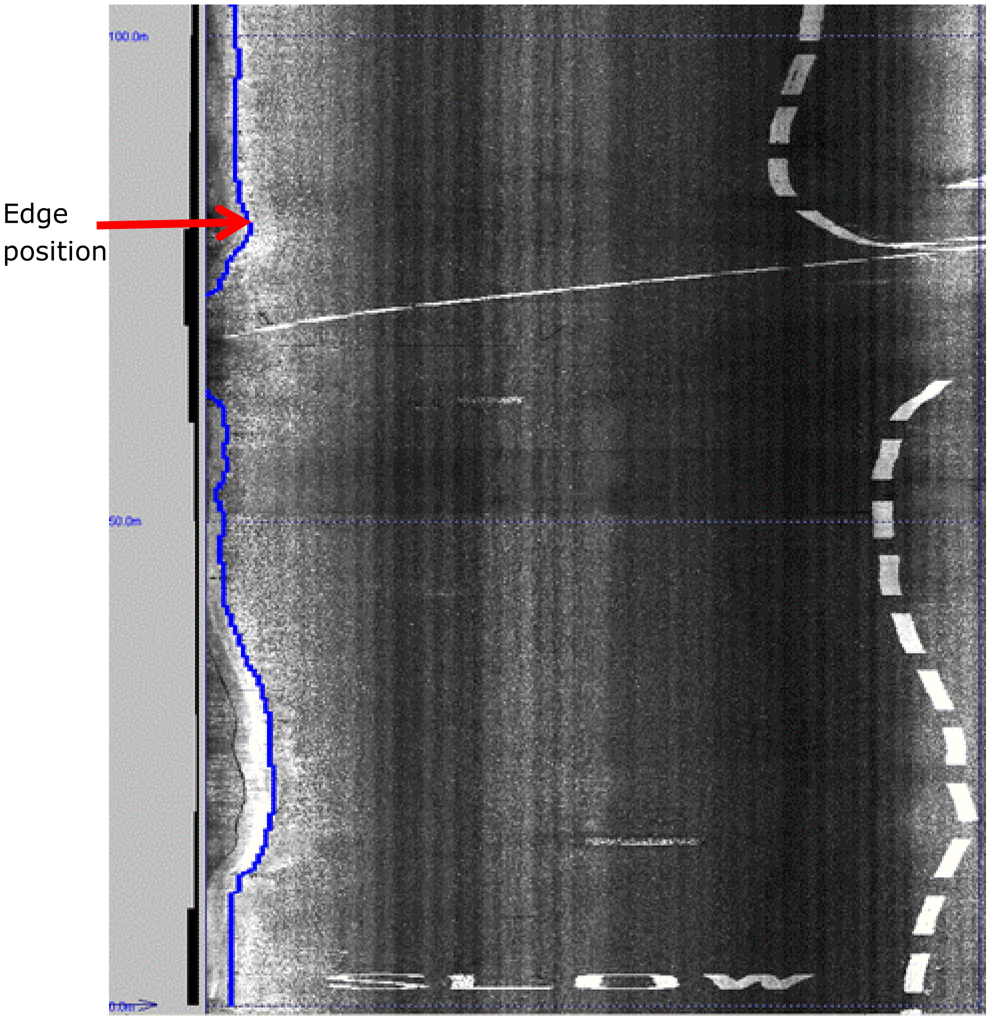



**Figure 42: Results of manual analysis to identify the road edge**

As well as comparing the edge position determined by the cleaned rut algorithm, we also compared the edge detection of the Highways England (TRACS) algorithm with the reference data. As can be seen from Figure 43, the TRACS edge position (red) is pretty good and matches the manual assessment (black) pretty well. However, the cleaned rutting edge position does not match well at all and is extremely different in places.

The fact that the cleaned rutting performs worse than the TRACS algorithm is surprising, as TRACS is very basic and either looks for a road marking or looks for unexpected changes in height, whereas the cleaned rutting algorithm is much more sophisticated.

On closer inspection, it was noted that the lengths where the edge has been incorrectly detected by the cleaned rutting algorithm are where issues with the consistency of the cleaned rutting are also seen. Thus it seems to be that one of the key things to obtaining an accurate and consistent measure of the rut depths on a road is for the edge to be detected well.

This analysis would also suggest that the current cleaned rutting algorithm would not improve the measurement of rutting, nor would it improve the consistency of the measure. Thus it has not been pursued further.

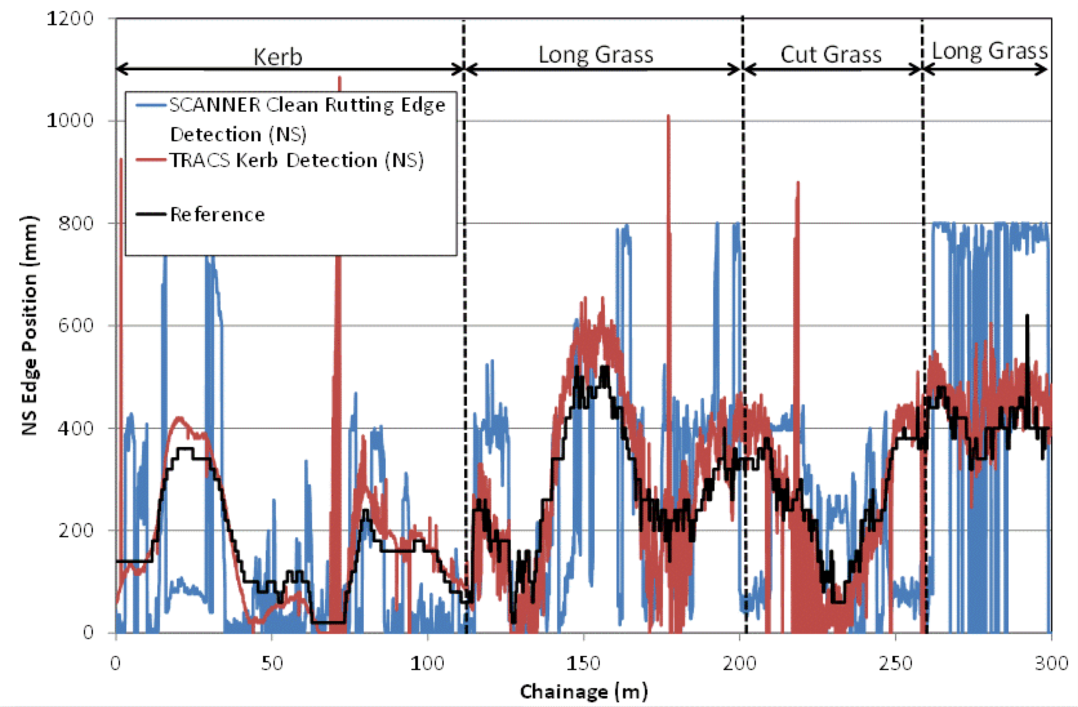



**Figure 43: Edge position, measured from left hand edge of transverse profile**

### *B.2.2 Use of high resolution data*

The repeatability of rutting data, when calculated from high resolution data can be compared to that calculated from low resolution data. This has been achieved by comparing the errors between repeat runs using HARRIS2 high resolution data on the SCANNER road routes, processed with the TRACS rutting algorithm, and rut depths provided by the SCANNER contractors. The cumulative frequency distribution of these errors is shown in Figure 44. As can be seen, the rut depths calculated from high resolution transverse profile (HARRIS2) are much more repeatable than those calculated from low resolution transverse profile (SCANNER) – there is a much higher percentage of differences <1mm for the high resolution data. This reflects the results seen when moving from 20 point transverse profiles in the TRACS2 contract, to 100 point profiles in the TRACS3 contract. It is worth noting that the nearside data is still less consistent than the offside.

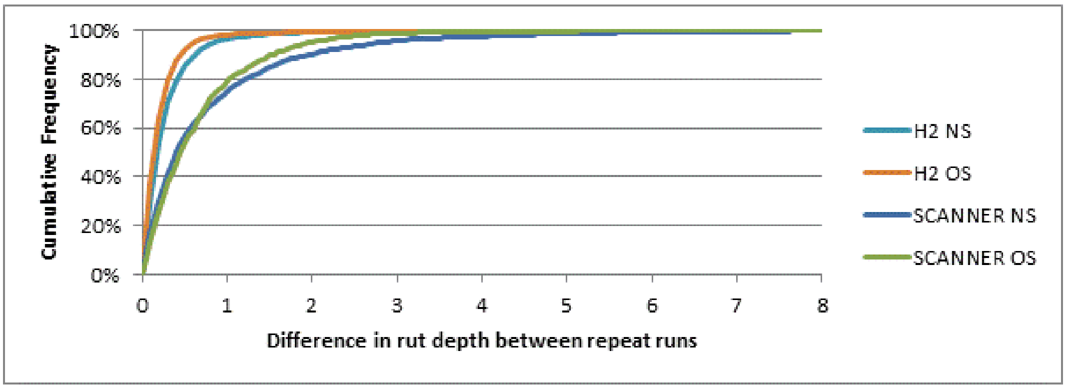



**Figure 44: Repeatability of rutting calculated from high resolution and low resolution transverse profile**

*B.2.2.1 Improved year on year consistency using high resolution data*

Since it was not possible to obtain high resolution profiles from the SCANNER contractors (raw data is not delivered by SCANNER), the improvements to repeatability that might be seen on the Principal roads have been investigated by assessing the repeatability of data from 2-way single carriageway roads on the trunk road network. Data for all 2 way single carriageway roads was extracted from Highway England's PMS (HAPMS) for surveys performed in 2010 and 2011 (low resolution system, equivalent to SCANNER) and then for surveys performed in 2013 and 2014 (high resolution system). The year on year repeatability was assessed for each type of system by calculating the difference between rut depths reports in 2010 and those reported in 2011 and then calculating the difference between 2013 and 2014. The results of this analysis are shown in Figure 45.


**Figure 45: Repeatability of TRACS data using low resolution (blue line) and high resolution (orange) transverse profile**

It can be seen, from Figure 45, that the frequency distribution of rut depth differences for the high resolution system (orange graph) has a much larger peak near 0 and is much narrower compared to the distribution from the low resolution system. Thus this would suggest that significant improvements in year on year consistency could also be obtained by using high resolution transvers profile data for the Principal roads.

The repeatability of SCANNER devices on non-principal roads, when assessed during the Accreditation tests, is pretty good already (Table 19) but it may be possible to improve this using high resolution data.

**Table 19: Percentage of lengths meeting 95% repeatability criteria for non-principal roads on SCANNER Road Route 2 (SRR2)**

| | SCANNER devices | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RAV5 | RAV6 | RAV7 | RAV8 | RAV9 | RAV10 | RAV11 | RAV12 | RAV14 | Tempest 1 | Tempest 2 | Tempest 3 |
| 2014 NS | 96% | 93% | 91% | 88% | 92% | 94% | 93% | 93% | | 93% | | |
| 2015 NS | 94% | 93% | 93% | 90% | 93% | 91% | 94% | 92% | 83% | 94% | 89% | 91% |
| **Average** | **95%** | **93%** | **92%** | **89%** | **92%** | **93%** | **93%** | **92%** | **83%** | **95%** | **89%** | **91%** |
| 2014 OS | 98% | 98% | 97% | 96% | 95% | 97% | 98% | 98% | | 94% | | |
| 2015 OS | 96% | 97% | 97% | 97% | 97% | 97% | 98% | 97% | 98% | 96% | 93% | 94% |
| **Average** | **97%** | **97%** | **97%** | **96%** | **96%** | **97%** | **98%** | **98%** | **98%** | **95%** | **93%** | **94%** |

To investigate this, the repeatability of data on SRR2 was assessed using high resolution HARRIS2 data and comparing this with the average performance of the SCANNER devices. If the cumulative frequency distribution, of differences between rut depth values for two repeat runs, is plotted it can be seen that the difference between rut depths, from two repeat runs on the non-principal roads in SRR2, are smaller in general for the high resolution data (HARRIS2 in Figure 46). Therefore it appears that, for the offside, using high resolution transverse profiles result in much more repeatable rut depths than when using low resolution data.

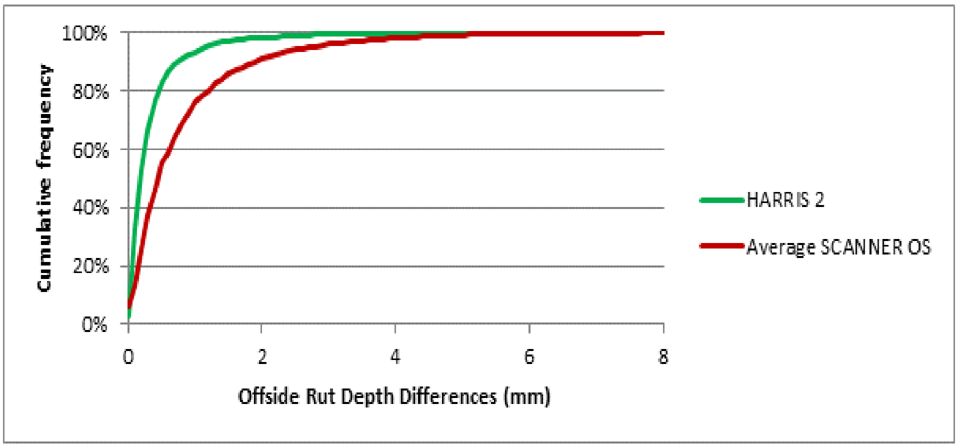



**Figure 46: Cumulative frequency of differences for offside rut depths from repeat runs**

However, when nearside data is considered, the two types of system perform similarly (Figure 47). Investigation, into why the performance of the nearside data isn't as good as for the offside, showed that repeatability is lower for the high resolution system where verges are present: If these are excluded, the repeatability of the high resolution system for nearside rutting is again better than the low resolution system (Figure 48).

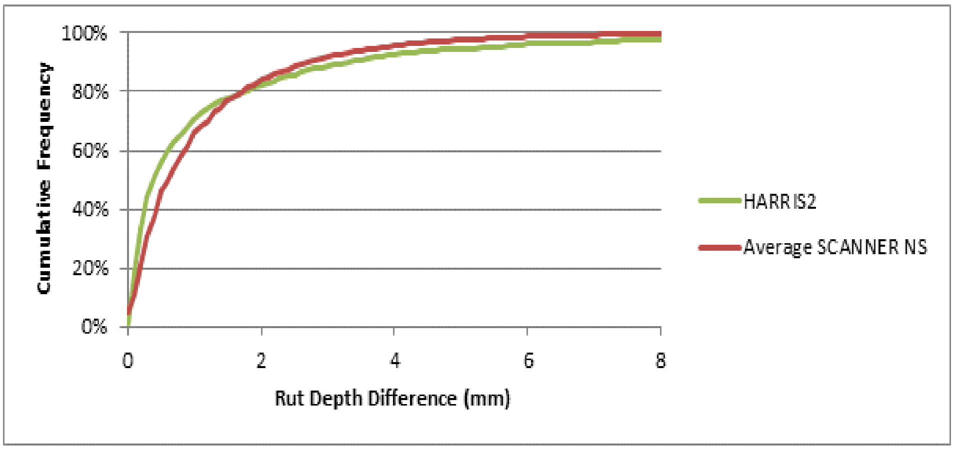



**Figure 47: Cumulative frequency of differences for nearside rut depths from repeat runs**

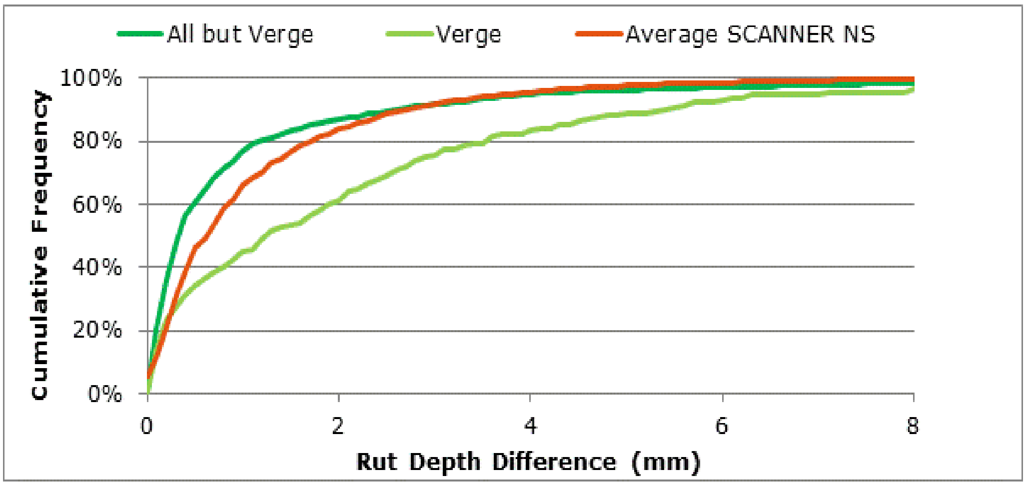



**Figure 48: Cumulative frequency of differences for repeat nearside rut depths: Green lines are rut depths calculated from high resolution data (HARRIS2)**

Thus for principal roads and non-principal roads, without verges, using a high resolution system and the TRACS rutting algorithm, provides much more repeatable data than using a low resolution system. Improving the TRACS rutting algorithm to cope better with verges is discussed further in Section B.2.3.

#### *B.2.2.2 Improved accuracy of rut depths calculated from high resolution data*

An investigation of rut depth accuracy was performed, to determine whether rut depths, calculated from high resolution transverse profiles, would be more accurate than those from low resolution data. There wasn't scope in the project to collect true reference data (i.e. with a straight edge and wedge) so a manual assessment of transverse profiles and forward facing images was used to determine the actual rut depths present on SRR2.

High resolution transverse profiles were obtained from a HARRIS2 survey of SRR2 and these were then re-sampled to generate low resolution transverse profiles, consisting of 22 points over a width of 3.2m (to replicate SCANNER). Both types of transverse profile were processed through the TRACS rutting algorithm, to eliminate any effects of using different algorithms on the results.

The comparison of nearside rut depths from high and low resolution data, with reference data is shown in Figure 49, whilst that for offside rut depths is given in Figure 50. As can be seen, the rut depths calculated from high resolution transverse profiles match the reference much better than those from low resolution data, particularly for the nearside.

This would suggest that using high resolution transverse profile data would improve the accuracy of the data provided, as well as the repeatability and the year on year consistency.

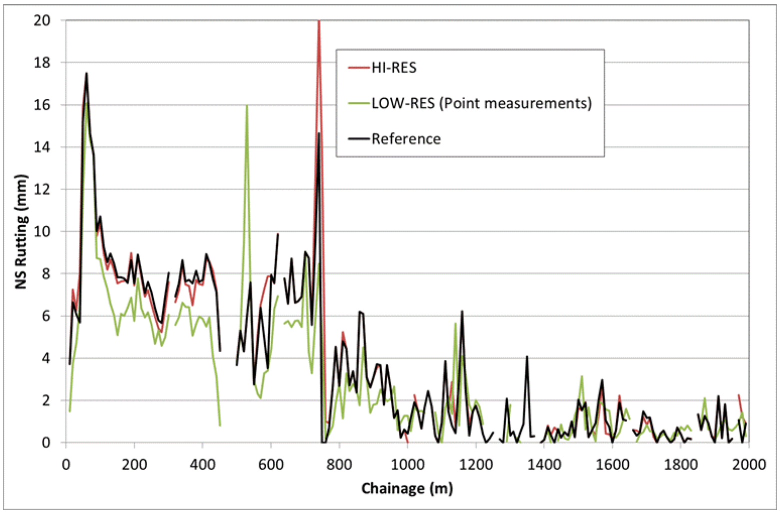



**Figure 49: Nearside rut depths, calculated from high and low resolution transverse profiles, compared with reference**

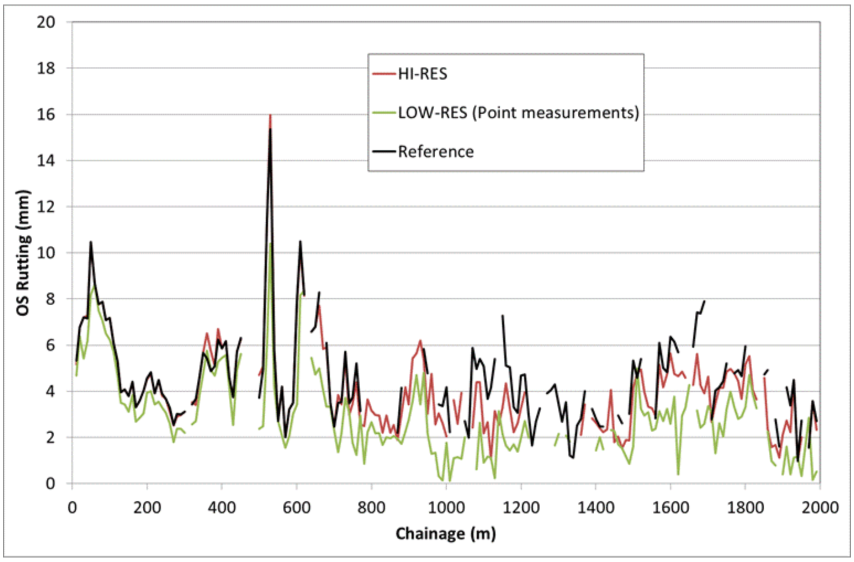



**Figure 50: Offside rut depths, calculated from high and low resolution transverse profiles, compared with reference**

### *B.2.3 Investigation of low class roads (verge/edge detection)*

It had been observed that the repeatability and accuracy of rut depths, provided by the TRACS rutting algorithm, was relatively poor when there were verges present at the edges of the road. It was not possible to investigate this further with the data already available: HARRIS2 data is only available for SRR1 and SRR2, and high resolution transverse profiles are not available as standard from SCANNER surveys. Therefore, a separate survey with HARRIS2 was commissioned to collect data from local U roads with a variety of road edges, including flat grass verges, hedges, raised verges.

The route was an extension to the current SRR2 route (Figure 51) to incorporate more lengths of challenging edges, such as that shown in Figure 52.

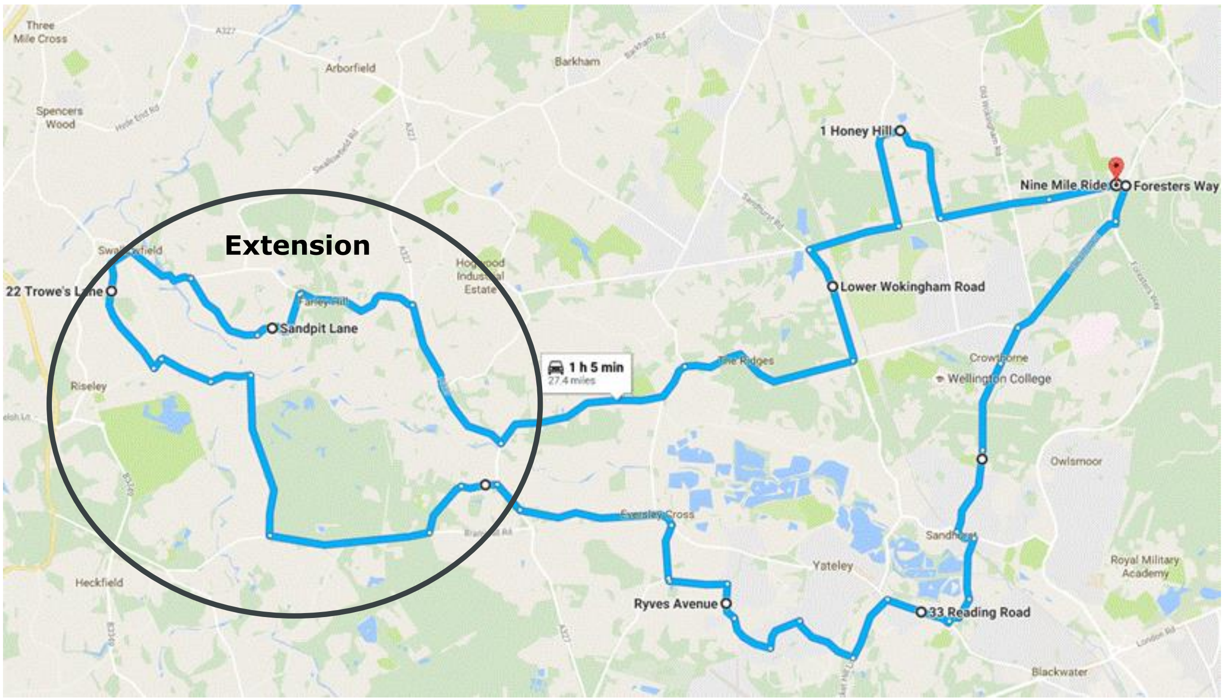



**Figure 51: Map showing SCANNER road route 2 and the proposed extension**

**Figure 52: Example of road edge found on the SRR2 extension**

*B.2.3.1 Edge detection*

The same method as used in Section B.2 was used to obtain reference data for the placement of the lane edge i.e. the downward and forward facing video was manually analysed to identify the edge of the lane. This reference edge position has been compared with the edge position obtained when processing the data with the TRACS rutting algorithm. This comparison is shown graphically in Figure 53 for the nearside and Figure 54 for the offside. Note that the offside position is measured as distance from the left hand edge of the images.

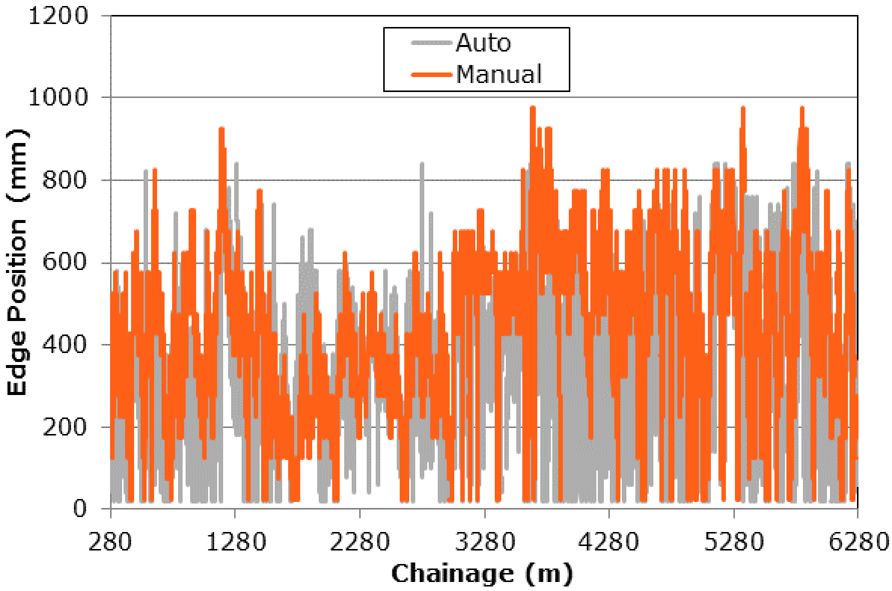



**Figure 53: Nearside edge position for 6km of the SRR2 extension**

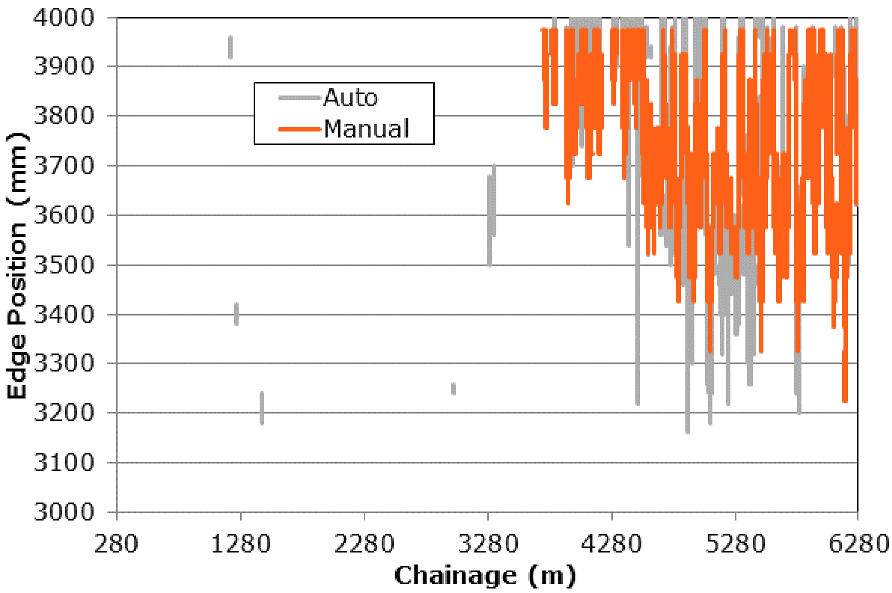



**Figure 54: Offside edge position for 6km of the SRR2 extension**

As can be seen from Figure 53 and Figure 54, the automatic edge detection is relatively noisy and contains gaps. However, in general it performs relatively well and, on some lengths, particularly well (Figure 55). The gaps are not necessarily a problem for the use of the output to invalidate the transverse profile as the TRACS rutting algorithm contains an algorithm that generates a smooth edge, even when broken road marking are present and the same approach might also be sufficient for the edge detection even where road markings are not present.

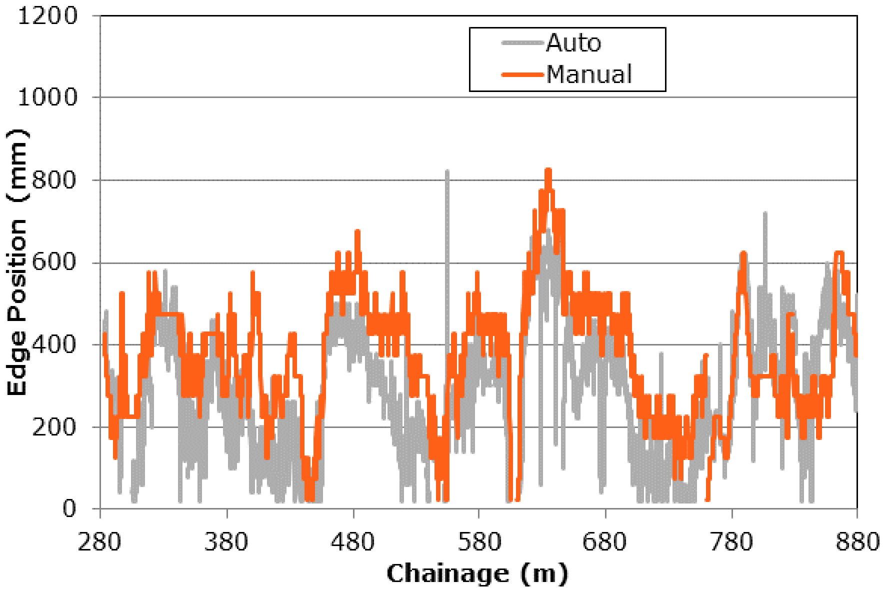



**Figure 55: Length where the automatic edge detection matches the reference well**

Whilst there is general agreement, there are some lengths where the automatic algorithm disagrees with the manual analysis, for example the lengths shown in Figure 56. Inspecting the downward facing images for this length, it can be see that the automatic analysis has only picked up half of the verge – where it is light enough to be mistaken for a road marking – see Figure 57.

If the transverse profiles are inspected for this length, it can be seen that the incorrect placement of the road by the TRACS algorithm would have an effect on the rutting calculated (Figure 58).

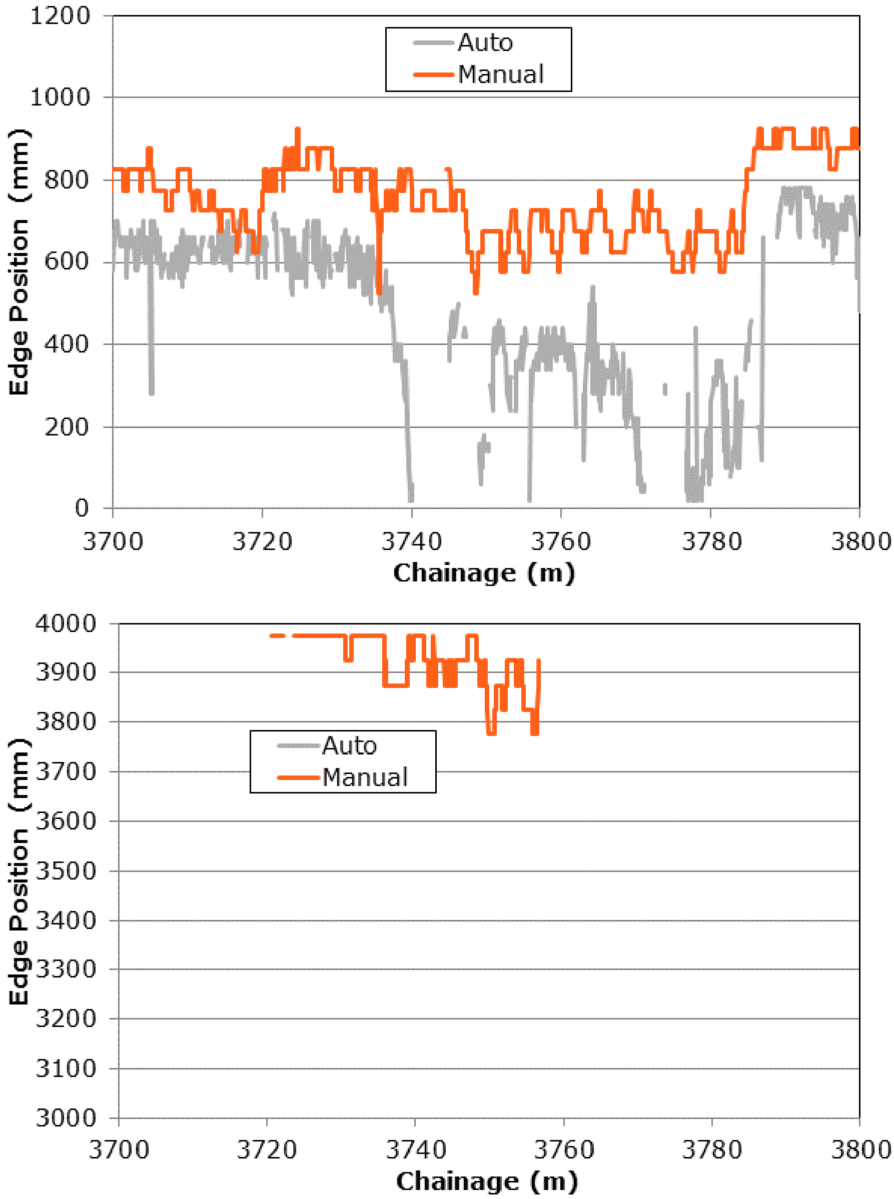



**Figure 56: An example of where the automatic edge placement disagrees with the reference: Nearside edge (top) and offside edge (bottom) for the same length**

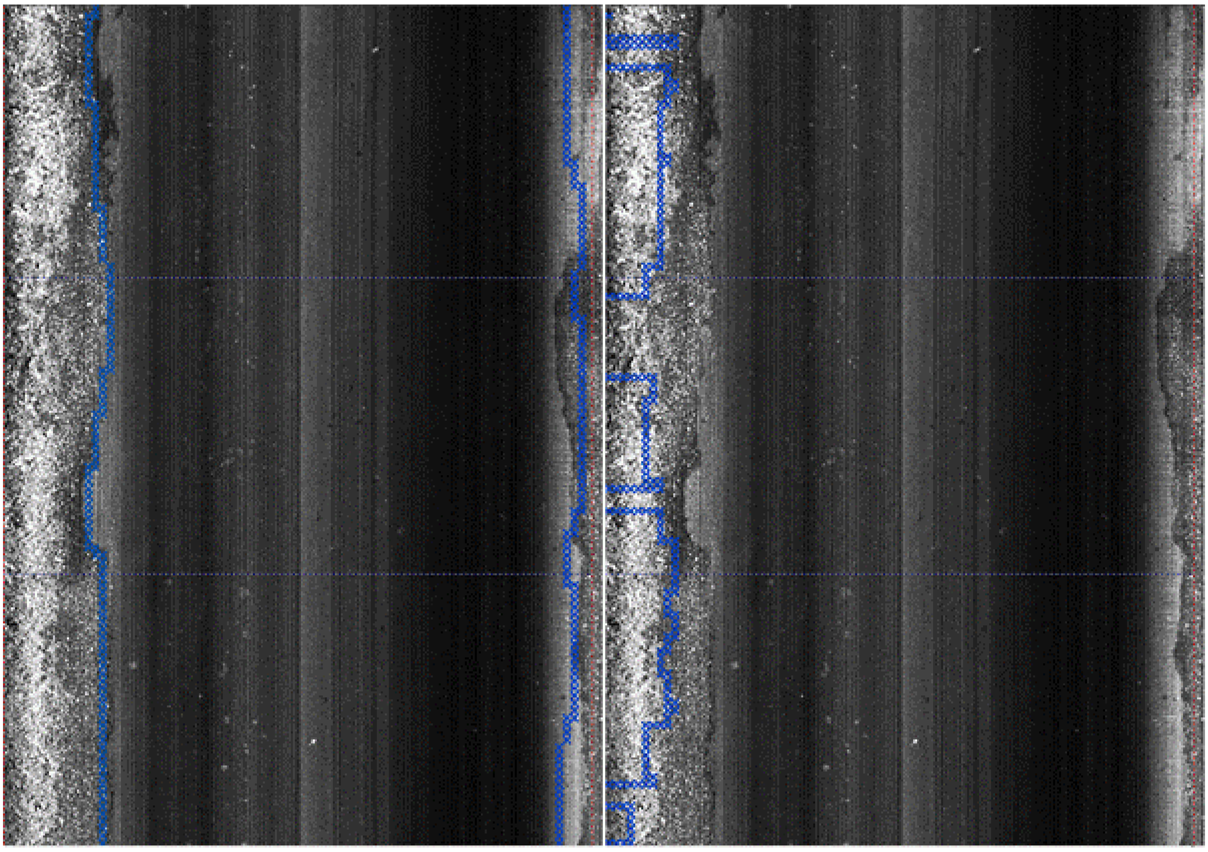


**Figure 57: Downward facing images from length where automatic edge detection disagrees with reference (images from ~3754m in Figure 56)**

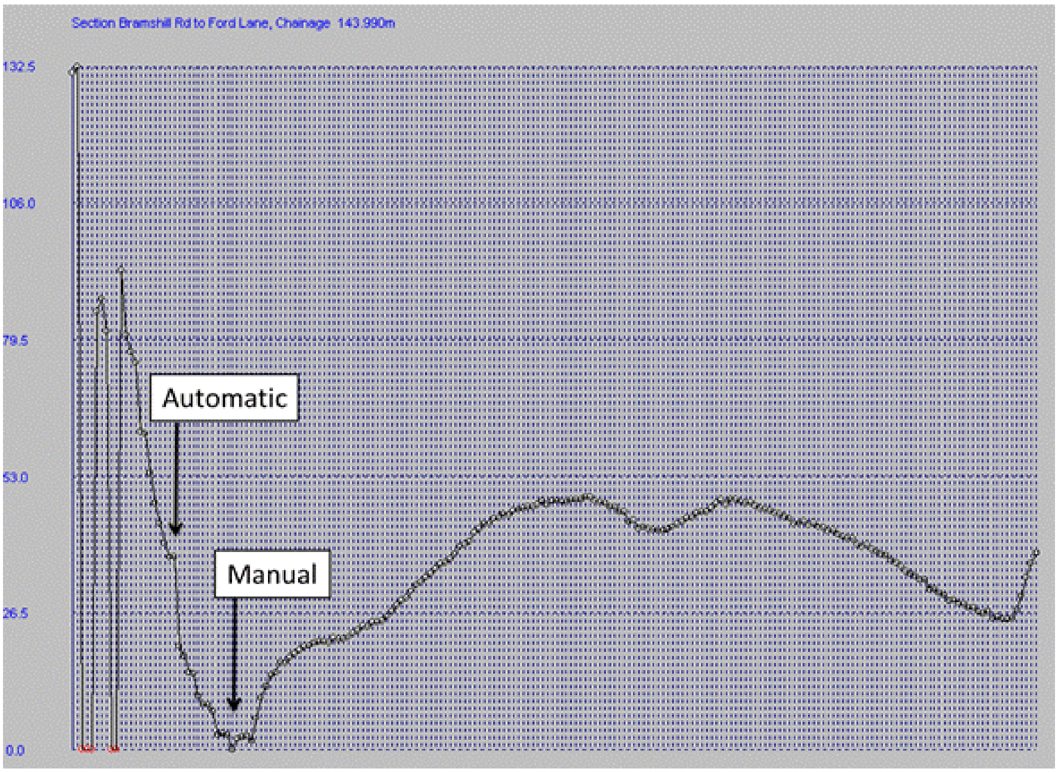



**Figure 58: Transverse profile from length of road shown in Figure 57 and placement of edge by TRACS rutting algorithm and reference (manual)**

### B.2.3.2 Rutting

Rut depths were calculated, using the TRACS algorithm for the SRR2 extension data. For some lengths, the algorithm calculated accurate and repeatable values but the following observations were made:

- There were an unusually large number of invalid rut values
- Even if the reference edge position was used in the rut depth calculation, this still resulted in many invalid rut values and poor repeatability of the data.

The transverse profiles and the placement of the straight edge were analysed further to determine why this might be and several reasons were found for this:

- Straight edge placed too close to the lane edge;
- Straight edges used to calculate nearside and offside ruts overlap;
- Straight edge for nearside rut placed in offside of profile;
- Transverse profile (consisting only of valid points) too narrow to calculate rut depths.

Examples of these are given below. All lengths have been chosen from the same length of road, for which the automatic edge position matched the reference well (Figure 55).

**Straight edge placed too close to the edge**

Figure 59 shows two consecutive transverse profiles and the placement of the straight edges used to calculate nearside rut (red line) and offside rut (blue line). For both profiles, the nearside edge has been correctly removed and the nearside straight edge placement is similar, resulting in a consistent nearside rut depth being calculated. However, in the first profile (top plot), at a chainage of 281m, the offside straight edge has been placed too far to the right when compared to the more realistic position shown for the subsequent profile (bottom plot). This has resulted in an inconsistent calculation of offside rut depth for these two lengths.

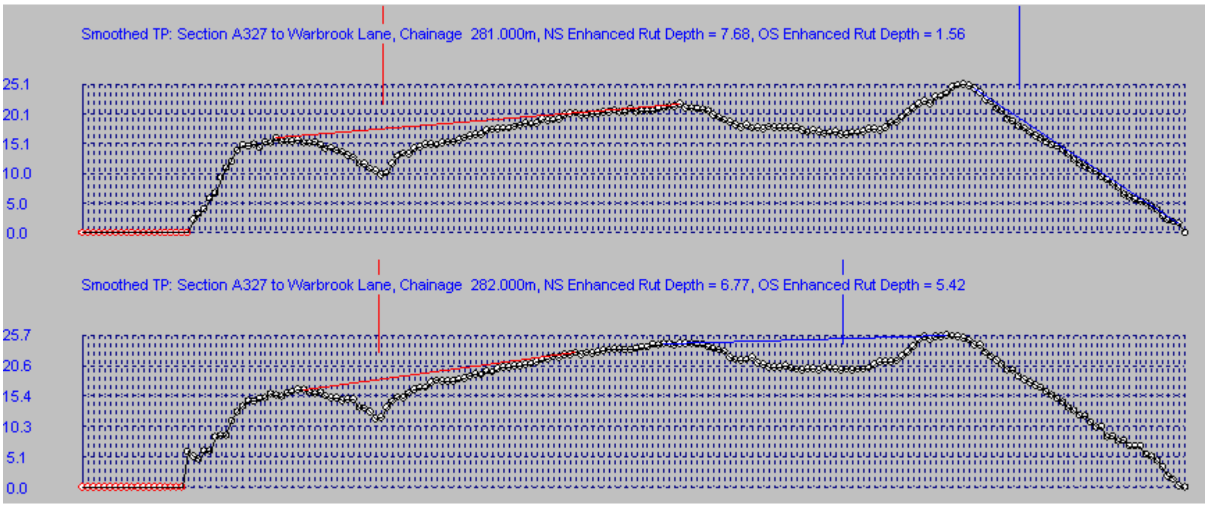



**Figure 59: An example of where the straight edge has been placed unrealistically close to the edge of the road (blue straight edge in top plot)**

**Overlapping straight edges**

Figure 60 shows an example of where the rutting algorithm has placed the nearside and offside straight edges so that they overlap, resulting in the rut depths being calculated in the same depression. This situation has probably arisen due to the narrowness of the valid part of the profile.

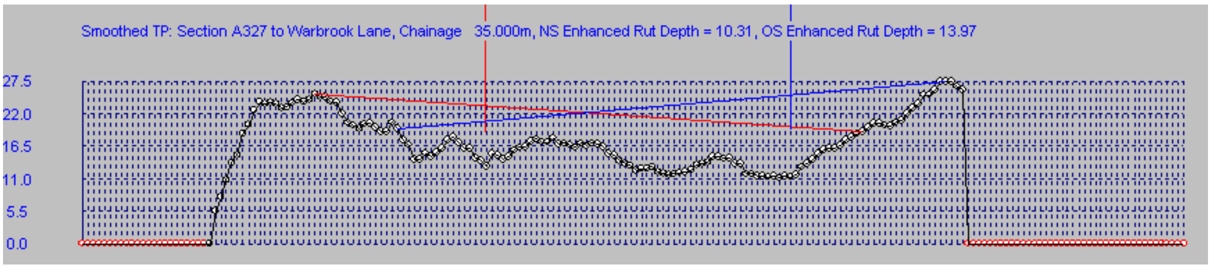



**Figure 60: An example of the rutting algorithm placing the straight edges so that they overlap**

**Nearside straight edge placed in offside**

Figure 61 shows a case where the rutting algorithm has placed the nearside straight edge over the offside of the road in the first survey run (top plot) but not found a rut in the nearside for the second survey (bottom plot). Similarly, in the first survey, the algorithm has not been able to place the straight edge in the offside but has in the second survey. This has resulted in apparently very different rut depths, despite the measured transverse profile being very similar to a human assessor.

The difference in straight edge placement for the two runs may be due to inconsistent identification of a road marking, being detected in the second survey (bottom plot) but not for the first survey. The reason for the inconsistency in the road marking detection was due to the marking being very worn (potentially the marking may have been previously removed/burnt off) – see Figure 62.

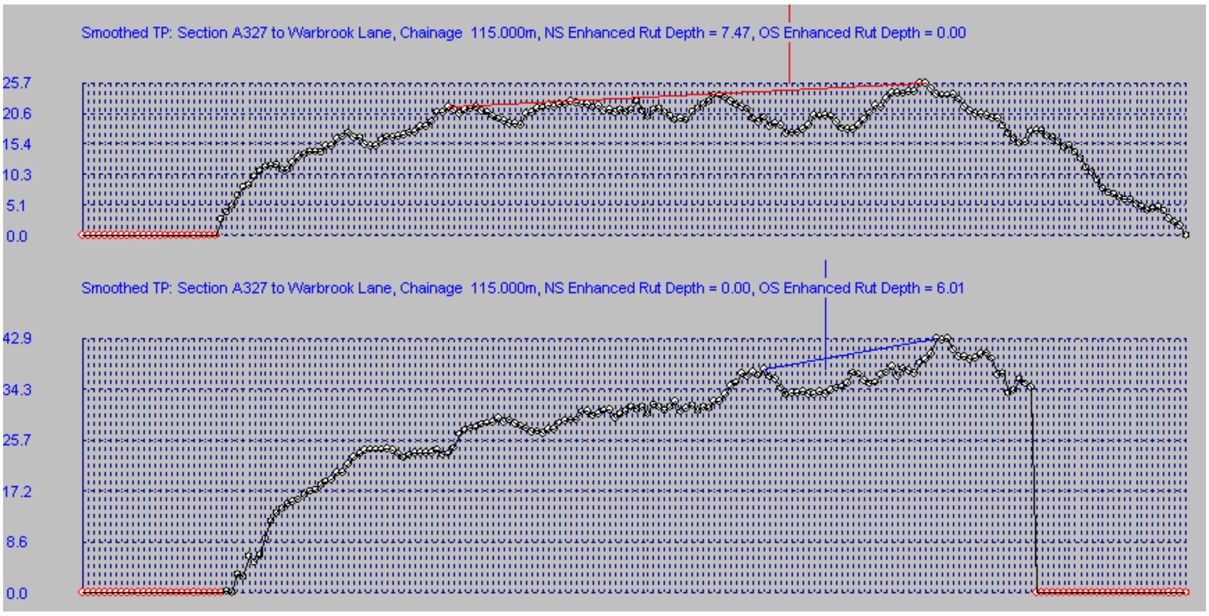



**Figure 61: An example of the rutting algorithm placing the NS rut in the OS wheelpath.**

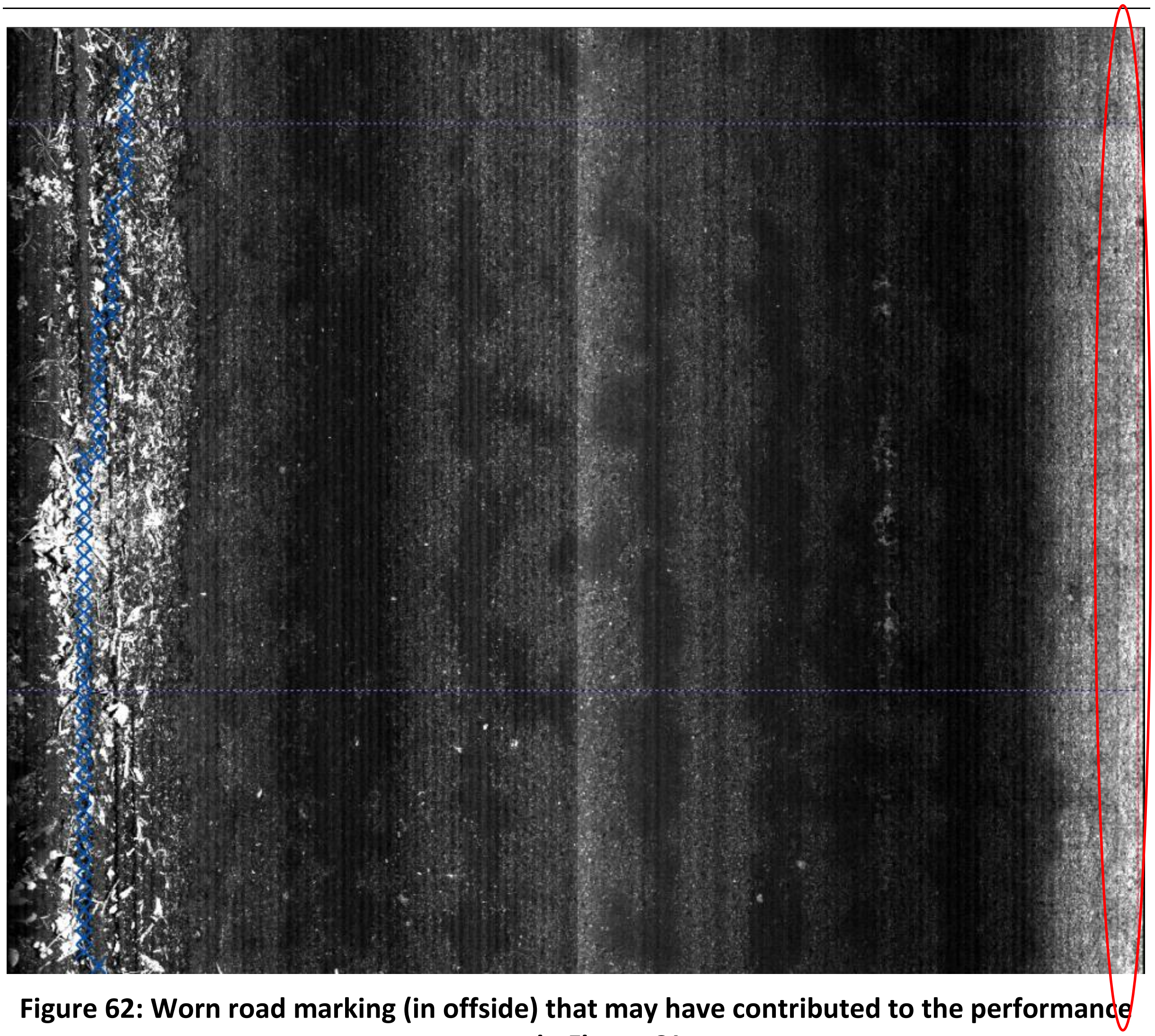


**Figure 62: Worn road marking (in offside) that may have contributed to the performance seen in Figure 61**

**Valid profile too narrow to calculate rutting**

The TRACS algorithm has a minimum width requirement, in order to be able to calculate a valid rut depth. The example shown in Figure 63 is of a length of road where the width of the valid profile was too narrow for the rutting algorithm.

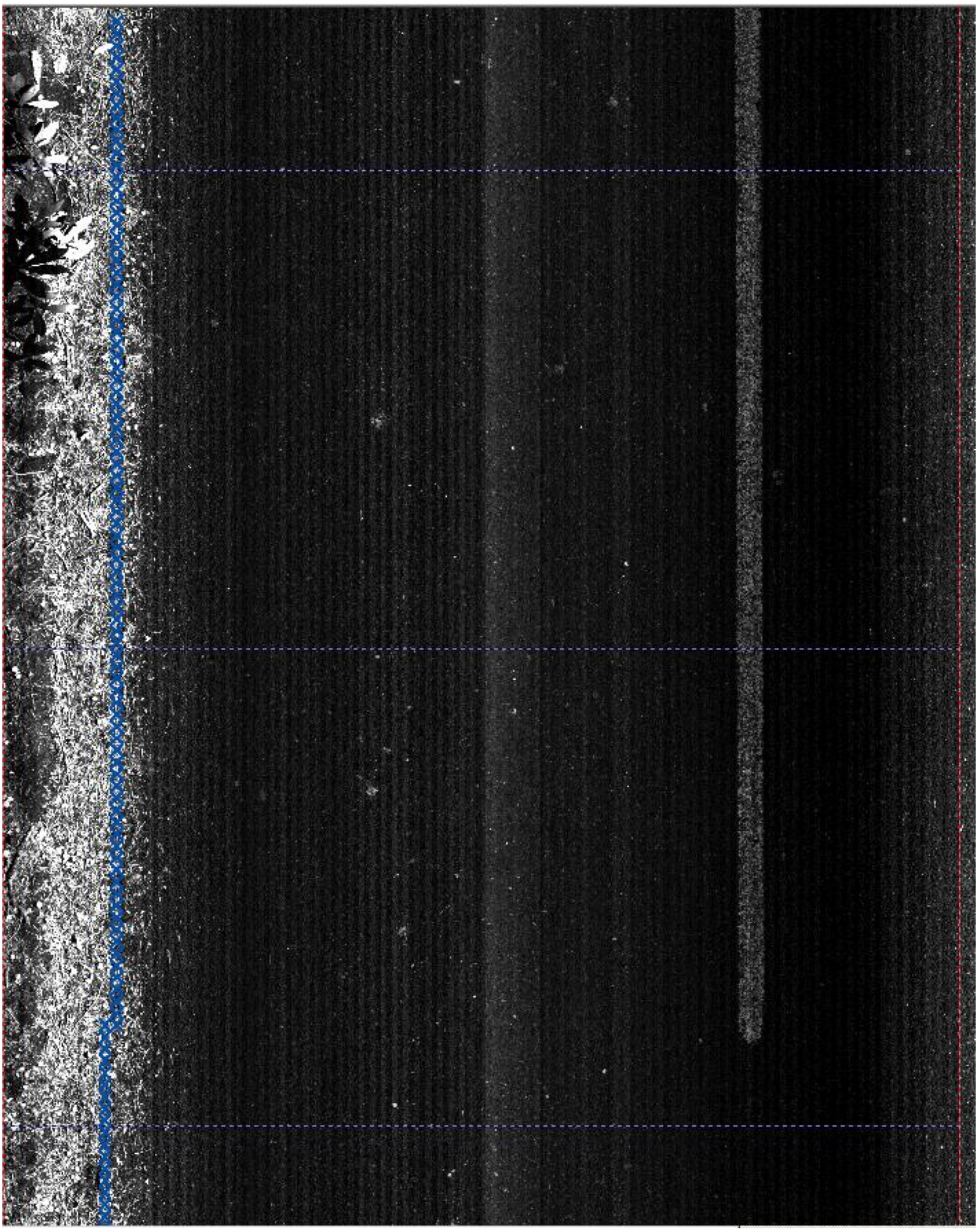


**Figure 63:An example of where the valid profile was too narrow for the rutting algorithm, due to the presence of a verge on the NS and a road marking on the OS (Warbrook Ln to St Neots Rd, 645m)**

#### B.2.4 Conclusion

The TRACS rutting algorithm appears to perform well on many examples of local roads. However, sometimes there are issues with edge detection and also with straight edge placement on the narrower, lower classed roads.

The following improvement to the automatic edge detection algorithm would be suggested:

- Require the edge position to smoothly vary down the road, to prevent the edge jumping about: This would be a relatively straightforward procedure to implement, as it could be based on the edge smoothing currently implemented for broken road markings.

There are also improvements to the straight edge placement that could be implemented, in order to make the rut depth calculation more consistent on the low class roads:

- A rut is a continuous feature and will not significantly shift transversally in adjacent profiles. Therefore the straight edge should be placed in a similar position in adjacent profiles, to ensure the same rut is measured for each profile;
- Require a minimum distance between the centre of the straight edges placed to calculate nearside and offside rut.

There are some lengths where the road is not wide enough to calculate valid rut depths (using any rutting algorithm). For these lengths:

- Users should be encouraged to use the transverse evenness parameter instead of rutting on very narrow roads.

# Appendix C SCANNER Condition Parameters (Task 2)

## C.1 Quick Wins - Ride Quality

### *C.1.1 Effect of geometry on LPV*

Research for TRACS showed that LPV values were affected by road geometry with high values being provided on bends and lengths where there is high crossfall, despite the ride quality on these lengths actually not being poor. More effect is seen on the parameters as the longer wavelengths are included, so 30m is worst affected and 3m least affected. Because of this, eLPV was developed and Highways England adopted this as their measure for ride quality in June 2004.

To determine the extent of the effect of geometry on LPV on the local road network, data from Devon, a large rural authority, has been used. Firstly, the behaviour of eLPV and LPV on the Devon network has been analysed by calculating the percentage of the network lying in each RCI category (Red Amber Green) for each parameter. This is shown in Table 20 and it can be seen, 3m eLPV places similar amounts of the network in each category to 3m LPV. However, 10m eLPV places more of the network in the Green category than LPV and less in Amber and Red.

Note that eLPV is not currently included in the RCI calculation, so the thresholds applied to TRACS eLPV and those previously applied to TRACS LPV data have been used to determine thresholds for eLPV that are equivalent to those used to calculate the RCI contribution from LPV. The thresholds used are given in Table 21.

**Table 20: Categorisation of Devon network by eLPV and LPV**

| | A roads (%) | | | B roads (%) | | | C roads (%) | | | All roads (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | G | A | R | G | A | R | G | A | R | G | A | R |
| **3m LPV** | 22.8 | 1.1 | 0.3 | 14.7 | 1.2 | 0.2 | 48.2 | 9.4 | 2.1 | 85.7 | 11.7 | 2.6 |
| **3m eLPV** | 22.9 | 1.0 | 0.3 | 14.8 | 1.1 | 0.2 | 50.8 | 7.2 | 1.7 | 88.5 | 9.3 | 2.2 |
| **10m LPV** | 19.9 | 3.2 | 1.1 | 12.3 | 2.8 | 1.0 | 33.4 | 14.3 | 9.1 | 65.7 | 23.2 | 11.2 |
| **10m eLPV** | 22.8 | 1.1 | 0.3 | 14.7 | 1.1 | 0.3 | 45.7 | 8.4 | 2.7 | 83.3 | 13.6 | 3.2 |

**Table 21: Thresholds applied to eLPV, equivalent to those used for LPV in the RCI calculation**

| Parameter | Road class | Lower threshold | Upper threshold |
|---|---|---|---|
| 3m eLPV | A | 2.2 | 5.5 |
| | B | 2.7 | 7.1 |
| | C | 3.8 | 9.3 |
| | U | 4.4 | 10.9 |
| 10 eLPV | A | 8.5 | 22.8 |
| | B | 11 | 28.8 |
| | C | 14.2 | 37.7 |
| | U | 16.6 | 44.6 |

The RCI category for each length of road, as determined by each parameter has also been calculated and whether the categories from the two measures agree determined (Table 22).

**Table 22: Comparison of RCI categories for eLPV and LPV, split by road type, for Devon**

| | | LPV | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A roads (%) | | | B roads (%) | | | C roads (%) | | | All roads (%) | | |
| | | G | A | R | G | A | R | G | A | R | G | A | R |
| 3m eLPV | G | 22.6 | 0.3 | 0.0 | 14.5 | 0.3 | 0.0 | 47.3 | 3.4 | 0.1 | 84.4 | 4.0 | 0.1 |
| | A | 0.2 | 0.7 | 0.1 | 0.2 | 0.9 | 0.0 | 0.9 | 5.7 | 0.6 | 1.3 | 7.3 | 0.7 |
| | R | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.2 | 0.0 | 0.3 | 1.4 | 0.0 | 0.4 | 1.8 |
| 10m eLPV | G | 19.7 | 2.5 | 0.6 | 12.1 | 2.1 | 0.5 | 31.9 | 10.7 | 3.1 | 63.8 | 15.3 | 4.2 |
| | A | 0.2 | 0.6 | 0.3 | 0.2 | 0.6 | 0.3 | 1.5 | 3.1 | 3.8 | 1.9 | 7.3 | 4.4 |
| | R | 0.0 | 0.1 | 0.2 | 0.0 | 0.1 | 0.2 | 0.0 | 0.5 | 2.2 | 0.0 | 0.6 | 2.6 |

The results shown in Table 22 suggest that 19.7% of lengths have been classed as green by both 10m eLPV and 10m LPV, 2.5% have been classed as green by eLPV but Amber by LPV and 0.6% classed as green by eLPV but as Red by LPV.

If the two parameters behaved identically, it would be expected that all values off the main diagonal (i.e. non-grey squares) would be zero. It might also be expected that the values on either side of the diagonal would be roughly equal, if the two parameters behaved similarly. However, this is not generally the case, suggesting that the parameters do not behave similarly for all road classes.

Experience of the LPV parameter in TRACS showed that the LPV parameter was affected by road geometry, with unexpectedly high values being obtained for lengths with large gradients or small radii of curvature. Thus it would be expected that more of the network would be classed as poor by the LPV parameter. Looking at the results for all road classes in Table 22 (right hand columns) we can see that this is the case for both 3m LPV and 10m LPV – values to the right of the grey squares are larger than values to the left of the squares, significantly so for the 10m parameters.

Looking at only A roads or B roads, 3m eLPV and 3m LPV are equivalent: 0.4% of the network is given a worse condition category by 3m LPV, whereas 0.3% of the network is given a worse condition category by 3m eLPV. Thus 3m LPV and eLPV can be considered to be equivalent on A and B roads.

However, there is some difference between the 10m LPV and 10m eLPV parameters, with more of the network categorised in a worse condition by the LPV parameter. This can be expected, as there will be some lengths on A and B roads where the geometry might be artificially increasing the LPV values.

The difference between the parameters is most apparent on the C roads, with a significant amount more lengths placed in the higher categories for both 3m and 10m LPV. Again this could be expected, as we'd expect the more extreme geometries to be present on the lower classes of road, and for there to be more curvature in general.

To confirm that the differences are due to geometry each length of the network was classed by gradient, curvature and crossfall, with 5 classes developed for each characteristic. How these classes have been defined is given in Table 23.

**Table 23: Categories for Road geometry**

| Category | Curvature | Crossfall | Gradient |
|---|---|---|---|
| **1** | Straight: > 1000m | Adequate crossfall: less than 2% | Flat: less than 2% |
| **2** | Moderately curved: between 500m and 1000m | Slightly higher than adequate crossfall: between 2% and 3% | Low gradient: between 2% and 5% |
| **3** | Fairly curved: between 100m and 500m | Moderately high crossfall: between 3% and 7% | Moderately steep: between 5% and 10% |
| **4** | Curved: between 50m and 100m | High crossfall: between 7% and 10% | Steep: between 10% and 15% |
| **5** | Extremely curved: <50m. | Very high crossfall: greater than 10% | Very steep: greater than 10% |

When looking at the lengths of road in the lower classes of geometry (classes 1-3), the two parameters match fairly well (Table 24 and Table 25). However, for the higher classes (4 and 5), there is usually the same amount or a larger amount of the network classified as red by 3m LPV than 3m eLPV, especially for Curvature. For example, 3m LPV reports twice as many lengths in the Red category as 3m eLPV for class 5 Curvature on A roads (0.06% compared with 0.03%) and slightly more in the Amber category (0.17% and 0.11%).

For the higher geometry classes significantly more of the network is classed as Red by 10m LPV than 10m eLPV (Table 25).

This would suggest that the LPV parameter values are artificially raised by high levels of road geometry on local roads as well as on the trunk road network.

**Table 24: Percentage of network lying in each RCI category (Green, Amber, Red) for 3m LPV and 3m eLPV, split by road geometry class**

| Road class | Geometry class | Curvature | | | | | | Crossfall | | | | | | Gradient | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 3m LPV | | | 3m eLPV | | | 3m LPV | | | 3m eLPV | | | 3m LPV | | | 3m eLPV | | |
| | | G | A | R | G | A | R | G | A | R | G | A | R | G | A | R | G | A | R |
| A | 1 | 48.7 | 1.62 | 0.46 | 48.7 | 1.51 | 0.47 | 30.0 | 1.46 | 0.39 | 30.2 | 1.24 | 0.38 | 39.1 | 1.56 | 0.45 | 39.2 | 1.42 | 0.45 |
| | 2 | 16.3 | 0.78 | 0.20 | 16.4 | 0.67 | 0.23 | 25.6 | 0.97 | 0.27 | 25.7 | 0.87 | 0.28 | 33.9 | 1.46 | 0.42 | 34.1 | 1.27 | 0.43 |
| | 3 | 25.9 | 1.31 | 0.50 | 26.1 | 1.44 | 0.48 | 36.1 | 1.86 | 0.57 | 36.2 | 1.71 | 0.56 | 19.1 | 1.20 | 0.35 | 19.3 | 1.04 | 0.33 |
| | 4 | 2.69 | 0.29 | 0.07 | 2.77 | 0.21 | 0.07 | 2.44 | 0.14 | 0.06 | 2.48 | 0.11 | 0.04 | 2.05 | 0.22 | 0.06 | 2.07 | 0.20 | 0.06 |
| | 5 | 0.72 | 0.17 | 0.06 | 0.81 | 0.11 | 0.03 | 0.16 | 0.02 | 0.01 | 0.17 | 0.01 | 0.00 | 0.13 | 0.01 | 0.01 | 0.13 | 0.01 | 0.00 |
| B | 1 | 39.2 | 2.43 | 0.45 | 39.3 | 2.33 | 0.45 | 34.1 | 2.73 | 0.58 | 34.3 | 2.58 | 0.53 | 35.4 | 2.77 | 0.61 | 35.6 | 2.61 | 0.59 |
| | 2 | 15.8 | 1.28 | 0.26 | 15.9 | 1.23 | 0.27 | 20.8 | 1.50 | 0.32 | 20.9 | 1.4 | 0.32 | 31.0 | 2.50 | 0.49 | 31.3 | 2.26 | 0.44 |
| | 3 | 31.1 | 2.88 | 0.6 | 31.4 | 2.6 | 0.58 | 34.0 | 2.97 | 0.62 | 34.4 | 2.65 | 0.6 | 20.8 | 1.83 | 0.40 | 21.1 | 1.61 | 0.39 |
| | 4 | 3.58 | 0.57 | 0.14 | 3.67 | 0.49 | 0.13 | 1.91 | 0.31 | 0.06 | 1.95 | 0.27 | 0.07 | 3.25 | 0.37 | 0.08 | 3.26 | 0.36 | 0.08 |
| | 5 | 1.11 | 0.37 | 0.13 | 1.28 | 0.26 | 0.07 | 0.09 | 0.01 | 0.00 | 0.09 | 0.01 | 0.00 | 0.39 | 0.06 | 0.01 | 0.39 | 0.07 | 0.01 |
| C | 1 | 31.0 | 4.19 | 0.70 | 32.0 | 3.26 | 0.63 | 40.4 | 7.82 | 1.63 | 42.6 | 6.03 | 1.29 | 27.5 | 4.08 | 0.90 | 28.5 | 3.27 | 0.76 |
| | 2 | 15.0 | 2.58 | 0.48 | 15.6 | 2.03 | 0.42 | 16.8 | 2.95 | 0.63 | 17.5 | 2.32 | 0.51 | 27.5 | 4.58 | 0.94 | 28.8 | 3.47 | 0.74 |
| | 3 | 29.0 | 6.44 | 1.41 | 30.7 | 5.00 | 1.12 | 22.8 | 4.61 | 1.12 | 24.1 | 3.58 | 0.86 | 20.0 | 4.56 | 1.01 | 21.4 | 3.40 | 0.75 |
| | 4 | 4.13 | 1.41 | 0.40 | 4.59 | 1.06 | 0.29 | 0.90 | 0.24 | 0.09 | 0.98 | 0.19 | 0.06 | 5.21 | 1.94 | 0.49 | 5.71 | 1.57 | 0.36 |
| | 5 | 1.72 | 1.01 | 0.48 | 2.16 | 0.78 | 0.27 | 0.03 | 0.01 | 0.01 | 0.04 | 0.01 | 0.00 | 0.73 | 0.47 | 0.15 | 0.80 | 0.43 | 0.12 |

**Table 25: Percentage of network lying in each RCI category (Green, Amber, Red) for 10m LPV and 10m eLPV, split by road geometry class**

| Road class | Geometry class | Curvature | | | | | | Crossfall | | | | | | Gradient | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 10m LPV | | | 10m eLPV | | | 10m LPV | | | 10m eLPV | | | 10m LPV | | | 10m eLPV | | |
| | | G | A | R | G | A | R | G | A | R | G | A | R | G | A | R | G | A | R |
| A | 1 | 44.7 | 4.90 | 1.08 | 48.8 | 1.60 | 0.33 | 26.0 | 4.37 | 1.43 | 30.0 | 1.53 | 0.28 | 35.4 | 4.47 | 1.24 | 39.2 | 1.59 | 0.34 |
| | 2 | 14.5 | 2.23 | 0.63 | 16.4 | 0.78 | 0.14 | 23.1 | 2.86 | 0.85 | 25.7 | 0.94 | 0.20 | 29.6 | 4.70 | 1.45 | 34.0 | 1.49 | 0.30 |
| | 3 | 21.2 | 4.94 | 1.87 | 26.0 | 1.65 | 0.38 | 31.2 | 5.46 | 1.82 | 36.2 | 1.82 | 0.44 | 15.8 | 3.46 | 1.42 | 19.2 | 1.19 | 0.27 |
| | 4 | 1.70 | 0.85 | 0.50 | 2.70 | 0.28 | 0.07 | 1.89 | 0.46 | 0.3 | 2.45 | 0.16 | 0.03 | 1.47 | 0.54 | 0.32 | 2.09 | 0.18 | 0.05 |
| | 5 | 0.29 | 0.28 | 0.38 | 0.75 | 0.16 | 0.05 | 0.07 | 0.06 | 0.07 | 0.17 | 0.02 | 0.00 | 0.08 | 0.04 | 0.03 | 0.13 | 0.01 | 0.00 |
| B | 1 | 34.7 | 6.00 | 1.37 | 39.5 | 2.25 | 0.36 | 28.8 | 6.49 | 2.14 | 34.4 | 2.53 | 0.47 | 31.3 | 5.75 | 1.75 | 35.6 | 2.63 | 0.52 |
| | 2 | 13.7 | 2.84 | 0.82 | 16.0 | 1.16 | 0.21 | 17.9 | 3.57 | 1.15 | 20.9 | 1.46 | 0.28 | 26.1 | 5.88 | 2.01 | 31.2 | 2.34 | 0.45 |
| | 3 | 25.3 | 6.80 | 2.51 | 31.3 | 2.73 | 0.57 | 28.4 | 6.71 | 2.47 | 34.3 | 2.75 | 0.61 | 16.6 | 4.58 | 1.88 | 21.0 | 1.71 | 0.36 |
| | 4 | 2.26 | 1.23 | 0.80 | 3.60 | 0.55 | 0.14 | 1.36 | 0.56 | 0.36 | 1.93 | 0.28 | 0.08 | 2.28 | 0.97 | 0.45 | 3.32 | 0.31 | 0.08 |
| | 5 | 0.49 | 0.47 | 0.65 | 1.13 | 0.34 | 0.15 | 0.04 | 0.03 | 0.04 | 0.09 | 0.01 | 0.00 | 0.24 | 0.17 | 0.06 | 0.41 | 0.05 | 0.01 |
| C | 1 | 23.5 | 9.17 | 3.26 | 29.5 | 5.47 | 0.97 | 27.6 | 14.9 | 7.38 | 38.1 | 9.63 | 2.16 | 21.4 | 7.96 | 3.15 | 26.3 | 5.14 | 1.09 |
| | 2 | 10.7 | 5.25 | 2.16 | 14.2 | 3.26 | 0.65 | 12.0 | 5.60 | 2.74 | 15.9 | 3.61 | 0.81 | 19.0 | 9.56 | 4.42 | 25.9 | 5.86 | 1.26 |
| | 3 | 19.0 | 11.6 | 6.26 | 27.3 | 7.70 | 1.87 | 15.8 | 7.99 | 4.71 | 21.6 | 5.42 | 1.45 | 12.3 | 8.15 | 5.08 | 18.8 | 5.41 | 1.37 |
| | 4 | 2.20 | 1.93 | 1.81 | 3.91 | 1.52 | 0.51 | 0.57 | 0.34 | 0.32 | 0.85 | 0.28 | 0.11 | 2.85 | 2.70 | 2.09 | 4.89 | 2.10 | 0.65 |
| | 5 | 0.64 | 0.91 | 1.67 | 1.67 | 1.00 | 0.55 | 0.01 | 0.02 | 0.03 | 0.03 | 0.01 | 0.01 | 0.44 | 0.47 | 0.43 | 0.73 | 0.44 | 0.18 |

### C.1.2 *Consistency of eLPV on SCANNER road routes*

Data from each year's Accreditation tests is used to calculate the consistency of each RCI parameter and this data is used to provide the "expected range" applied to each parameter within the QA audits. To determine the range that we would expect the parameter to vary by each year, the bias and random error for each machine is calculated. These values have been calculated for eLPV, using data from the last 3 years, and compared with the values for LPV (calculated already for the QA audits). The values for bias and random error for 3m eLPV are plotted in Figure 64 and those for the 10m parameters are similar. It can be seem from these graphs that 3m eLPV bias and random error are significantly smaller than those for LPV, so it might be concluded at this point that eLPV is significantly more consistent than LPV. However, this isn't a very fair comparison, as LPV values are larger and have a wider range than eLPV values.

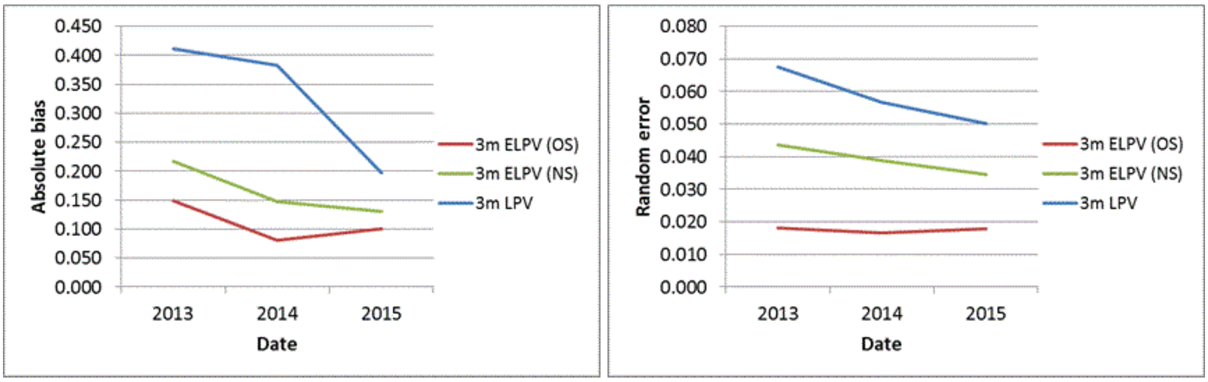



**Figure 64: Bias (left) and Random Error (right), calculated for 3m LPV and 3m eLPV**

So, to account for this difference in the measures, the data has been normalised, to make the eLPV "equivalent" to LPV. This has been achieved by taking the range of LPV values for this data and the equivalent for eLPV and applying the ratio of these to the eLPV data. The results of this are plotted in Figure 65. Note that a direct comparison can only be performed between the blue and green lines on the graph, as these are both for profile measured in the nearside; the red line (offside profile) has just been included for information.

When eLPV is considered in this way, it can be seen that eLPV still has smaller bias and random errors than LPV in general or has a very similar bias and random error. Thus it can be concluded that eLPV is as consistent, or is more consistent than LPV on the SCANNER road routes.

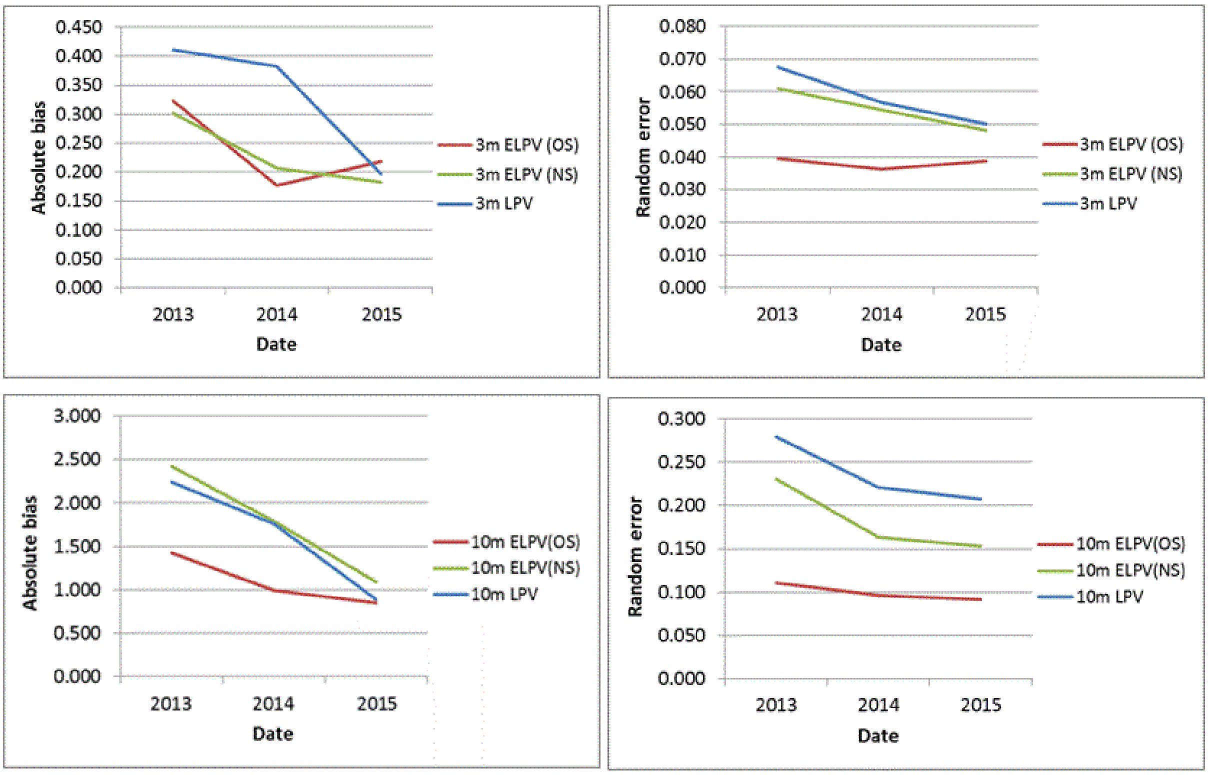



**Figure 65: Bias (left) and Random Error (right), calculated for LPV and scaled eLPV: 3m parameters (top) and 10m parameters (bottom)**

### *C.1.3 Consistency of eLPV for QA audits*

eLPV has been shown to be as consistent as or more consistent than LPV on the SCANNER road routes. There is now a need to investigate whether using it on the network would help with consistency there. So, two authorities where the QA Audit reports had shown there was an issue with LPV consistency were identified: Devon and Herefordshire. eLPV was used instead of LPV for these authorities, to determine whether using eLPV would improve the consistency seen, particularly in the RCI and Audit Indicator.

The QA Audit report for Devon is shown in Figure 66 and as can be seen, the bias from the previous year for 10m LPV is outside of the expected range, being almost twice the upper limit of the expected range. The amount of lengths contributing the maximum amount (Red) to the RCI is also inconsistent for this parameter. The Audit Indicator has increased from 5.0% in 2013/14 to 13.9% in 2015/16, well outside of the expected range.

The Audit report has been reproduced but replacing LPV with eLPV and this is shown in Figure 67. New expected ranges for the bias for the eLPV parameters have been calculated, using Accreditation data. As can be seen, the bias for 10m eLPV now lies within the expected range and the amount of lengths contributing the maximum to the RCI is more similar for the two year. However, whilst the difference between the Audit Indicators for the two years is now slightly smaller, the Audit Indicator calculated for 2015/16 is still much larger than would be expected. This suggests that the difference in the Audit Indicator was not down to LPV alone.

The situation is similar for Herefordshire, with the bias between years for 10m LPV being out of the expected range (Figure 68). When eLPV is used instead, the bias falls within the expected range (Figure 69).

| Statistics | | | | | | | RCI | | |
|---|---|---|---|---|---|---|---|---|---|
| Param. | Data | Mean | 99th percentile[1] | Distn. Corr. | Bias from Prev. | Expect Bias Range | Red[3] | Amber[3] | Green[3] |
| Nearside Rut | Curr. | 5.65 | 17.60 | 0.99 | 0.64 | -0.68 to 0.68 | 0.4% | 12.0% | 87.6% |
| | Prev. | 5.01 | 14.30 | | | | 0.1% | 6.7% | 93.2% |
| Offside Rut | Curr. | 5.44 | 17.30 | 0.97 | 1.16 | -0.52 to 0.52 | 0.4% | 11.5% | 88.1% |
| | Prev. | 4.29 | 13.70 | | | | 0.1% | 5.1% | 94.8% |
| Nearside 3m LPV | Curr. | 4.07 | 35.10 | 1.00 | 0.87 | -1.24 to 1.24 | 7.7% | 22.7% | 69.6% |
| | Prev. | 3.21 | 26.90 | | | | 4.6% | 17.0% | 78.4% |
| Nearside 10m LPV | Curr. | 32.91 | 611.29 | 1.00 | 4.57 | -2.26 to 2.26 | 22.9% | 25.1% | 52.0% |
| | Prev. | 28.35 | 438.60 | | | | 16.3% | 23.8% | 59.9% |
| Nearside SMTD | Curr. | 0.68 | 0.21 | 0.97 | -0.08 | -0.07 to 0.07 | 6.1% | 36.2% | 57.7% |
| | Prev. | 0.77 | 0.24 | | | | 3.5% | 28.3% | 68.2% |
| Cracking | Curr. | 0.34 | 3.80 | 0.99 | 0.16 | -0.17 to 0.17 | 5.1% | 29.7% | 65.2% |
| | Prev. | 0.18 | 2.20 | | | | 1.6% | 22.3% | 76.1% |
| RCI | Curr. | 49.09 | 180.32 | 0.99 | 16.70 | -12.29 to 12.29 | 13.9% | 37.3% | 48.7% |
| | Prev. | 32.39 | 135.00 | | | | 5.0% | 30.8% | 64.3% |

| Audit Indicator[4] | Curr. | 13.9% | Expected Range of Indicator for Curr. year |
|---|---|---|---|
| | Prev. | 5.0% | 3.5% to 6.5% |

[1] 1st Percentile for NS SMTD [2] Using RCI thresholds for urban B roads [4] Percentage of RCI values ≥ 100 (**Red** Category)
[3] Red = Max contribution to the RCI, Amber = Contributing to the RCI but not maximum, Green = Not contributing at all

**1 of 3 audit criteria met**

**Notes & Suggested Actions:**
Offside Rut, Nearside 10m LPV, Nearside SMTD and RCI are all outside expected bias from previous survey
The Audit Indicator is greater than expected for the current year

For further information please contact your current SCANNER survey provider.
If you then require additional information please email scanner@trl.co.uk.

**Figure 66: QA Audit report for Devon: Comparison of 2013/14 and 2015/16 data**

| Statistics | | | | | | | RCI | | |
|---|---|---|---|---|---|---|---|---|---|
| Param. | Data | Mean | 99th percentile[1] | Distn. Corr. | Bias from Prev. | Expect Bias Range | Red[3] | Amber[3] | Green[3] |
| Nearside Rut | Curr. | 5.64 | 17.40 | 0.99 | 0.63 | -0.68 to 0.68 | 0.4% | 12.0% | 87.6% |
| | Prev. | 5.01 | 14.30 | | | | 0.1% | 6.7% | 93.2% |
| Offside Rut | Curr. | 5.44 | 17.20 | 0.97 | 1.15 | -0.52 to 0.52 | 0.4% | 11.4% | 88.2% |
| | Prev. | 4.29 | 13.70 | | | | 0.1% | 5.2% | 94.8% |
| Nearside 3m LPV | Curr. | 1.89 | 17.57 | 1.00 | 0.37 | -1.23 to 1.23 | 5.4% | 19.4% | 75.2% |
| | Prev. | 1.51 | 13.22 | | | | 3.4% | 14.4% | 82.3% |
| Nearside 10m LPV | Curr. | 9.42 | 87.99 | 1.00 | 1.98 | -2.2 to 2.20 | 10.0% | 21.9% | 68.0% |
| | Prev. | 7.44 | 67.88 | | | | 6.1% | 17.1% | 76.7% |
| Nearside SMTD | Curr. | 0.68 | 0.21 | 0.97 | -0.08 | -0.07 to 0.07 | 6.1% | 36.1% | 57.7% |
| | Prev. | 0.77 | 0.24 | | | | 3.5% | 28.4% | 68.1% |
| Cracking | Curr. | 0.34 | 3.80 | 0.99 | 0.16 | -0.17 to 0.17 | 5.1% | 29.7% | 65.2% |
| | Prev. | 0.18 | 2.20 | | | | 1.6% | 22.3% | 76.1% |
| RCI | Curr. | 40.45 | 168.11 | 0.99 | 15.32 | -12.29 to 12.29 | 9.2% | 33.6% | 57.2% |
| | Prev. | 25.13 | 124.80 | | | | 2.9% | 23.7% | 73.4% |

| Audit Indicator[4] | Curr. | 9.2% | Expected Range of Indicator for Curr. year |
|---|---|---|---|
| | Prev. | 2.9% | 1.6% to 4.1% |

[1] 1st Percentile for NS SMTD [2] Using RCI thresholds for urban B roads [4] Percentage of RCI values ≥ 100 (**Red** Category)
[3] Red = Max contribution to the RCI, Amber = Contributing to the RCI but not maximum, Green = Not contributing at all

**1 of 3 audit criteria met**

**Notes & Suggested Actions:**
Offside Rut, Nearside SMTD and RCI are all outside expected bias from previous survey

The Audit Indicator is greater than expected for the current year

For further information please contact your current SCANNER survey provider.
If you then require additional information please email scanner@trl.co.uk.

**Figure 67: QA Audit report for Devon, using eLPV instead of LPV**

| Statistics | | | | | | | RCI | | |
|---|---|---|---|---|---|---|---|---|---|
| Param. | Data | Mean | 99th percentile[1] | Distn. Corr. | Bias from Prev. | Expect Bias Range | Red[3] | Amber[3] | Green[3] |
| Nearside Rut | Curr. | 5.61 | 16.10 | 0.96 | -0.60 | -0.71 to | 0.2% | 10.4% | 89.4% |
| | Prev. | 6.20 | 15.50 | | | 0.71 | 0.1% | 11.7% | 88.1% |
| Offside Rut | Curr. | 5.10 | 14.80 | 0.99 | 0.17 | -0.54 to | 0.1% | 7.9% | 92.0% |
| | Prev. | 4.93 | 14.80 | | | 0.54 | 0.1% | 8.1% | 91.8% |
| Nearside 3m LPV | Curr. | 3.07 | 28.10 | 0.99 | -0.43 | -1.3 to | 4.6% | 14.1% | 81.3% |
| | Prev. | 3.50 | 31.80 | | | 1.30 | 5.8% | 16.6% | 77.6% |
| Nearside 10m LPV | Curr. | 27.28 | 342.87 | 0.99 | -3.81 | -2.5 to | 13.0% | 24.8% | 62.2% |
| | Prev. | 31.09 | 440.63 | | | 2.50 | 16.7% | 27.7% | 55.5% |
| Nearside SMTD | Curr. | 0.76 | 0.23 | 1.00 | -0.02 | -0.07 to | 4.2% | 29.6% | 66.3% |
| | Prev. | 0.79 | 0.24 | | | 0.07 | 3.2% | 27.7% | 69.0% |
| Cracking | Curr. | 0.31 | 3.00 | 1.00 | 0.01 | -0.18 to | 3.4% | 32.6% | 64.0% |
| | Prev. | 0.29 | 2.90 | | | 0.18 | 2.9% | 32.0% | 65.1% |
| RCI | Curr. | 39.88 | 150.00 | 1.00 | 1.10 | -12.88 to | 7.6% | 36.0% | 56.4% |
| | Prev. | 38.78 | 145.61 | | | 12.88 | 7.0% | 35.6% | 57.4% |

| Audit Indicator[4] | | | Expected Range of Indicator for Curr. year |
|---|---|---|---|
| | Curr. | 7.6% | |
| | Prev. | 7.0% | 5.3% to 8.7% |

[1] 1st Percentile for NS SMTD [2] Using RCI thresholds for urban B roads [4] Percentage of RCI values ≥ 100 (**Red** Category)
[3] Red = Max contribution to the RCI, Amber = Contributing to the RCI but not maximum, Green = Not contributing at all

**2 of 3 audit criteria met**

**Notes & Suggested Actions:**
Bias in 10m LPV outside expected range.

For further information please contact your current SCANNER survey provider.
If you then require additional information please email scanner@trl.co.uk.

**Figure 68: QA Audit report for Herefordshire: Comparison of 2013/14 and 2015/16 data**

| Statistics | | | | | | | RCI | | |
|---|---|---|---|---|---|---|---|---|---|
| Param. | Data | Mean | 99th percentile[1] | Distn. Corr. | Bias from Prev. | Expect Bias Range | Red[3] | Amber[3] | Green[3] |
| Nearside Rut | Curr. | 5.61 | 16.10 | 0.96 | -0.60 | -0.71 to | 0.2% | 10.4% | 89.4% |
| | Prev. | 6.20 | 15.50 | | | 0.71 | 0.1% | 11.7% | 88.1% |
| Offside Rut | Curr. | 5.10 | 14.80 | 0.99 | 0.17 | -0.54 to | 0.1% | 7.9% | 92.0% |
| | Prev. | 4.93 | 14.80 | | | 0.54 | 0.1% | 8.1% | 91.8% |
| Nearside 3m LPV | Curr. | 1.50 | 15.44 | 0.99 | -0.20 | -1.24 to | 4.1% | 12.4% | 83.5% |
| | Prev. | 1.70 | 17.54 | | | 1.24 | 5.2% | 14.3% | 80.6% |
| Nearside 10m LPV | Curr. | 6.81 | 57.87 | 0.99 | -1.06 | -2.2 to | 4.5% | 14.9% | 80.7% |
| | Prev. | 7.87 | 66.55 | | | 2.20 | 5.8% | 17.8% | 76.4% |
| Nearside SMTD | Curr. | 0.76 | 0.23 | 1.00 | -0.02 | -0.07 to | 4.2% | 29.6% | 66.3% |
| | Prev. | 0.79 | 0.24 | | | 0.07 | 3.2% | 27.7% | 69.0% |
| Cracking | Curr. | 0.31 | 3.00 | 1.00 | 0.01 | -0.18 to | 3.4% | 32.6% | 64.0% |
| | Prev. | 0.29 | 2.90 | | | 0.18 | 2.9% | 32.0% | 65.1% |
| RCI | Curr. | 34.47 | 142.80 | 1.00 | 1.90 | -12.88 to | 5.7% | 32.1% | 62.2% |
| | Prev. | 32.57 | 140.00 | | | 12.88 | 5.1% | 30.5% | 64.5% |

| Audit Indicator[4] | | | Expected Range of Indicator for Curr. year |
|---|---|---|---|
| | Curr. | 5.7% | |
| | Prev. | 5.1% | 3.5% to 6.6% |

[1] 1st Percentile for NS SMTD [2] Using RCI thresholds for urban B roads [4] Percentage of RCI values ≥ 100 (**Red** Category)
[3] Red = Max contribution to the RCI, Amber = Contributing to the RCI but not maximum, Green = Not contributing at all

**All 3 audit criteria met**

**Notes & Suggested Actions:**

For further information please contact your current SCANNER survey provider.
If you then require additional information please email scanner@trl.co.uk.

**Figure 69: QA Audit report for Herefordshire, using eLPV instead of LPV**

## C.2 Quick Wins - Bump Measure

A small investigation has been carried out into why the Bump Measure is so inconsistent and whether it is practical to update it to provide a more consistent measure. This has been achieved by analysis of a network for which there is data from both 2013 and 2015 for the same lengths (Figure 70).

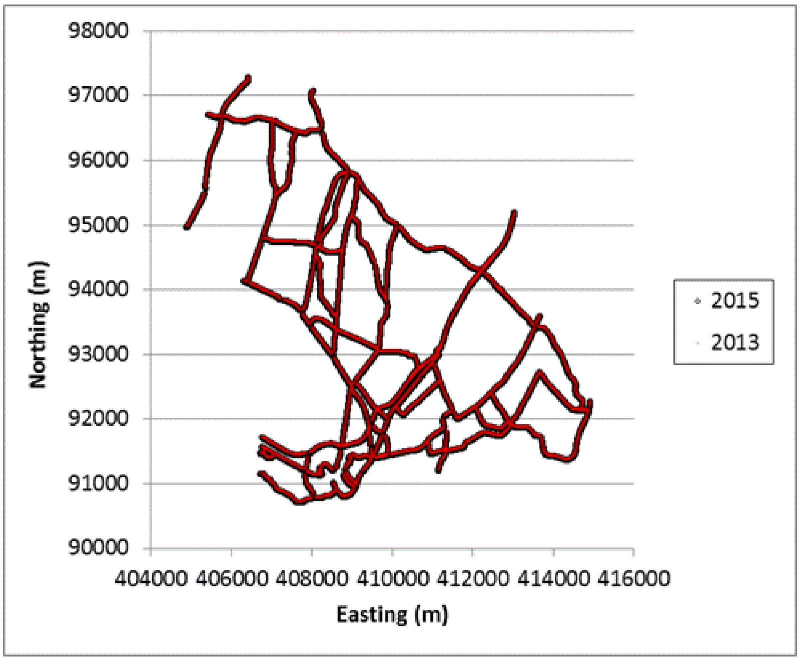



**Figure 70: Plot of the eastings and northings of data available for the Bournemouth network from 2013 and 2015**

Figure 71 shows the location of lengths containing a bump in the nearside in either 2013 or 2015. It can be observed that the number of bumps reported in 2013 is different from that in 2015: Approximately 0.9% of the network is affected in 2013, whereas in 2015 it is 0.5%. Even if a reporting length of 100m is used, this does not improve the consistency of bump reporting.

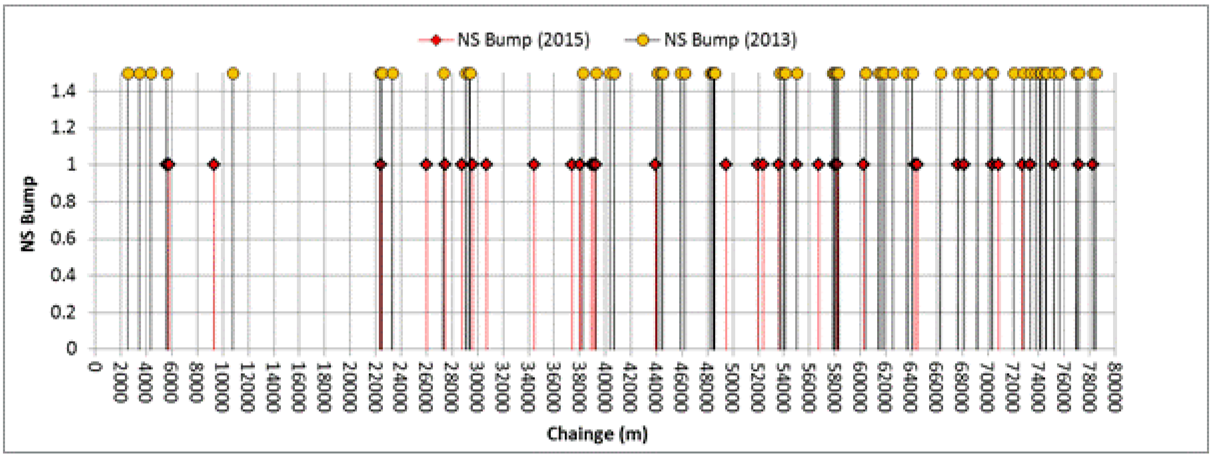



**Figure 71: Nearside bump reported in 2013 and 2015 for Bournemouth network**

If the lengths, where a bump has been reported, are inspected (using Google Streeview and Google Earth images), there is always a feature present that would cause users discomfort.

Figure 72 shows where bumps were reported on Boscombe Road – the yellow pins show lengths where no bump was identified in 2013, the blue pins where no bump was identified in 2015 and red pins showing where a bump was reported in either year. As can be seen, most lengths do not contain any bumps (one yellow, one blue pin) but there was one length reported to contain bumps in 2015 but not 2013 (one yellow, one red pin) and one length where a bump was reported both years (2 red pins).

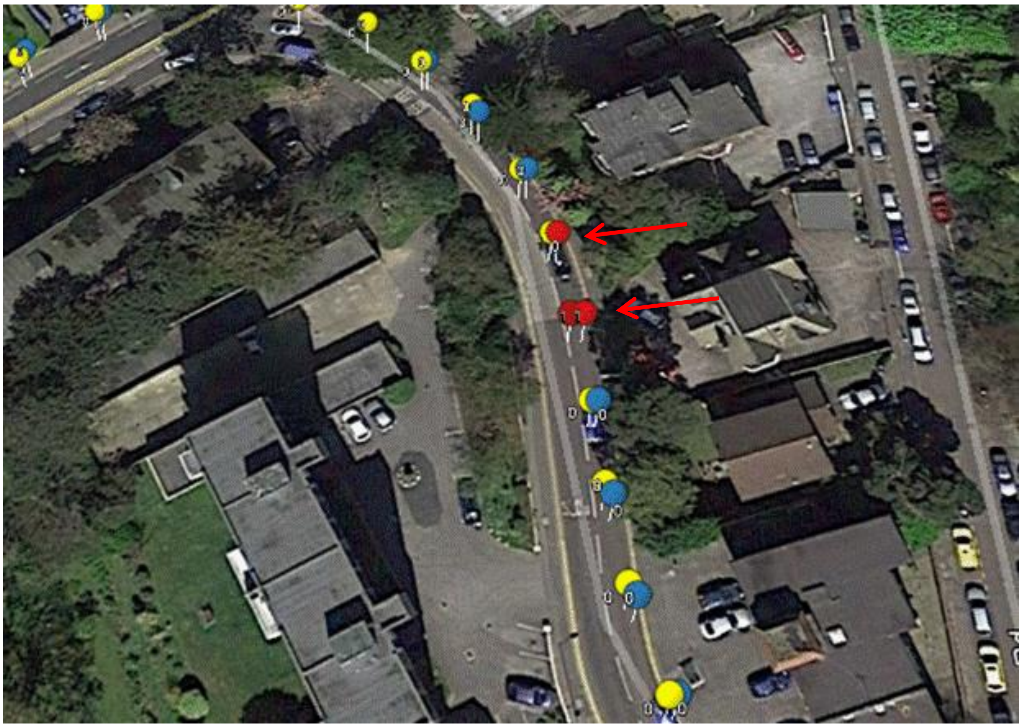


**Figure 72: Satellite view of Boscombe road in Bournemouth, showing where bumps have been identified in 2013 and 2015 (Google Earth)**

If these lengths are looked at in more detail, it can be seen that the length, where a bump is only reported in 2015, contains a failing pothole repair in the nearside wheelpath (Figure 73).The length where both surveys report a bump, contains a sunken trench (Figure 74).

**Figure 73: Failed pothole repair on Boscombe Road, reported to contain a bump in 2015 but not 2013**

**Figure 74: Sunken trench on Boscombe Road reported to contain a bump in 2013 and 2015**

So it seems that the bump measure does provide useful data but can be inconsistent as to whether a bump gets reported or not. This may be due to sensitivity to driving line, since it is calculated from a very thin longitudinal measurement line or may be due to the way that the parameter is calculated. It has not been possible, within the scope of the current project, to investigate this further.

# Appendix D Stakeholder Questionnaire used for Tasks 2 and 3

# Development of SCANNER condition surveys

**Consultation March 2016**

## D.1 The Development of SCANNER condition surveys consultation

The SCANNER survey on the local road network provides network wide condition assessment of the local A, B and C road network using survey devices that travel at traffic-speed measuring the shape of the road surface using laser sensors, and imaging the surface using digital cameras. The collected data is processed and converted into condition parameters, such as rutting, and delivered in a UKPMS compliant format to local authorities, for loading into their pavement management systems. The data is used within UKPMS compliant systems for the reporting of the SCANNER Road Condition Indicator (RCI) and the associated Highways Condition Index (HCI) figures for classified local authority roads in England and for the PIs used in Scotland, Wales and Northern Ireland. It is also used within these systems to identify lengths in need of maintenance or further investigation, to support scheme identification and prioritisation and to support asset valuation via the delivery of the Carriageway Condition Index (CCI), which is recognised by HAMFIG and CIPFA as an appropriate measure and methodology for use in Whole of Government Accounts (WGA).

In January 2015 the UKRLG commissioned TRL, supported by the Linhay and Hyperion consultancies, to undertake work to develop the SCANNER survey. One of the key objectives of this project is to determine how SCANNER could be improved to better meet the needs of local highway authorities in two areas:

- **The parameters delivered by SCANNER**. The SCANNER survey reports a wide range of condition parameters, covering road texture, ride quality, rutting, cracking and edge deterioration. Some of these were developed in research undertaken several years ago are not well used by authorities, or included in the SCANNER Road Condition Indicator (RCI). So, are the current parameters well used? Could any be rationalised, or removed? Are any measures missing?
- **The SCANNER Road Condition Indicator (RCI)**. Does the current method of reporting SCANNER data match how local highway authorities make maintenance decisions (or track the effects of maintenance undertaken)? Perhaps if this could be improved a stronger link could be developed between SCANNER data and maintenance activities.

To investigate these questions we are undertaking a consultation on the SCANNER survey and the data it provides, which will focus on these two areas.

## D.2 Consultation on the SCANNER Parameters

See the following table

| **Section D.2.1: Use of the parameters**<br>*We would like to better understand the use of the current core and enhanced SCANNER parameters and wish to seek views on the use, coverage, reliability, practicality, value and applications of these.*<br>*The following table presents the current list of SCANNER parameters. Could you provide a view on these parameters, using the following as a guide:*<br>• ***Current use**: Please describe whether and how you use this parameter currently e.g. to calculate a condition index, to identify or prioritise maintenance need on the network. State how often you use the parameter: e.g. frequent use, moderate use, little/no use.*<br>• ***Views**: Please give your views on this parameter, and your understanding of it, in engineering terms. Is it useful/valuable to you in managing the asset? Could it be more useful if improved in some way and what might this improvement be? Alternatively, you may feel that this is of little use. If so, why?*<br>• ***Importance rating**: If you were to consider this parameter in terms of its value to you in asset management, how would you score it in the range 1-5 (where 1 is very important and 5 is not important at all)? Please give reasons for your rating where appropriate.* | | | | |
|---|---|---|---|---|
| **Parameter** | **Core/ Enhanced (C/E)** | **My current use of this parameter** | **My views on this parameter** | **Importance rating (1-5), and reason** |
| **Road Roughness / shape** | | | | |
| 3m LPV (nearside, offside) | C | | | |
| 10m LPV (nearside, offside) | C | | | |
| Enhanced 3m LPV (nearside, offside) | E | | | |
| Enhanced 10m LPV (nearside, offside) | E | | | |
| Bump Measure (nearside, offside) | E | | | |
| Geometry (gradient, crossfall, curvature) | C | | | |
| **Rutting and transverse unevenness** | | | | |
| Rut Depths (nearside, offside) | C | | | |
| Cleaned Rut Depths (nearside, offside) | E | | | |
| Transverse variance | E | | | |
| Transverse unevenness (ADFD) | E | | | |

| | | | | |
|---|---|---|---|---|
| **Texture Parameters** | | | | |
| Texture (SMTD) | C | | | |
| Texture (MPD) | C | | | |
| RMST Texture depth in the nearside, centre and offside | E | | | |
| RMST Variance (nearside, centre, offside | E | | | |
| Texture Variability (RMST 5th Percentile, 95th Percentile, Variance) | E | | | |
| **Surface Deterioration Parameters** | | | | |
| Cracking (whole carriageway) | C | | | |
| Wheel Track Cracking (nearside, offside) | C | | | |
| Edge of carriageway cracking | C | | | |
| Other Visible Defect | C | | | |
| Transverse/reflection cracking | E | | | |
| Surface Deterioration | E | | | |
| **Edge Deterioration Parameters** | | | | |
| Edge roughness | E | | | |
| Edge steps (at two levels) | E | | | |
| Edge coverage | E | | | |

| | |
|---|---|
| **Section D.2.2: The enhanced parameters for rutting and variance**<br>*These are parameters that are a direct replacement for the current parameters*<br>• *Enhanced LPV vs Moving Average LPV*<br>• *Cleaned rutting vs Rutting* | |
| Do you use these enhanced parameters?<br>If not, please tell us why not | |
| If you do use them, please tell us:<br>Which do you use<br>Why – e.g. have you noted any benefits through applying these parameters in comparison with the originals? | |
| **Section D.2.3: Additional needs for SCANNER parameters**<br>*In this section we are seeking views on measures/parameters that are missing from SCANNER* | |
| Considering the list of parameters given above, what gaps do you see in the SCANNER data? E.g. defects on your road network, which you consider to be important, that SCANNER does not assess? Please list these and describe how you might use this information if SCANNER could be developed to provide this. | |

## D.3 Consultation on the SCANNER RCI

The aim of this section is to help us understand the extent to which authorities use the RCI, or use SCANNER parameters in another way, to inform maintenance decisions.

Using this information the SCANNER development project aims to consider how the RCI could be improved to provide more effective support for decisions about treatments.

| | |
|---|---|
| **Section D.3.1: Use of SCANNER data in maintenance decisions.**<br>*In this section we are examining how current use is made of the SCANNER in maintenance decisions* | |
| **Q1. Do you use SCANNER data to help you take decisions about maintenance?** | Yes / No |
| **Q2. If you answered *Yes* to Q1, which of the following do you use?** | |
| Treatments produced by UKPMS using the national treatment rules (i.e. UKPMS Rules & Parameters) | Yes/No *if Yes, please provide more detail* |
| An indicator (e.g. RCI, CCI, Edge CI or a locally-designed CI) | Yes/No *if Yes, please provide more detail* |
| SCANNER parameters directly | Yes/No *if Yes, please provide more detail* |
| Other | Yes/No *if Yes, please provide more detail* |

| **Q3. If you answered *No* to Q1, please explain your reasons** | |
|---|---|
| SCANNER parameters don’t give the type of information needed to make decisions about maintenance | Yes/No *if Yes, please provide more detail* |
| SCANNER parameters are appropriate, but aren’t collected reliably enough | Yes/No *if Yes, please provide more detail* |
| SCANNER parameters are not combined together in the right way in the UKPMS treatment rules and indicators | Yes/No *if Yes, please provide more detail* |
| We have other methods which are satisfactory | Yes/No *if Yes, please provide more detail* |
| Other | Yes/No *if Yes, please provide more detail* |

| **Section D.3.2: Use of SCANNER RCI.**<br>*In this section we are examining how current use is made of the SCANNER RCI, and views on how it may be improved or better linked to maintenance decisions* | |
|---|---|
| **Q4. Do you use the RCI to help you take decisions about maintenance?**<br>**Note: RCI is used in the calculation of national indicators such as 130-01 and 130-02 (England), SRMCS PI (Scotland), THS/012 (Wales). The RCI categorises lengths as red/amber/green.** | Yes/No *if Yes, please provide more detail* |
| **Q5. Do you use the RCI for any other purposes?** | Yes/No *if Yes, please provide more detail* |
| **Q6. Would you like closer links between the RCI and decisions about maintenance?** | Yes/No *please explain your answer* |
| **Q7. Would you be opposed to any changes being made to the RCI?** | Yes/No *please explain your answer* |
| **Q8. Do you have any suggestions for improvements to the RCI (e.g. its composition, weightings, etc.)?** | Yes/No *please provide more detail* |

## D.4 Background Information - What are the SCANNER Parameters and RCI?

SCANNER was developed from the Highways Agency's TRACS survey of the strategic road network. As TRACS was focussed fully on the measurement of roads that were well designed, typically wide, even and with few extremes of geometry, there was a need to undertake development of the survey to adopt it for local roads. A programme of research supported by the DfT was carried out between 2003 and 2007 to undertake this development. The primary outcomes of this work were revisions to the data collection requirements to better suit local roads and the delivery of parameters more focussed on narrower local roads, describing defects such as unevenness and edge deterioration. These have been applied, unchanged, since 2009 for network level SCANNER surveys.

There is a coordinate parameter (which consists of 3 attributes, X, Y, Z), used to locationally reference the data to the network, and 40 further parameters delivered by the SCANNER survey. These are listed in the following table, along with whether they were introduced in, or before, 2009.

All SCANNER parameters are reported at intervals of approximately 10m.

| UKPMS code | SCANNER survey parameter | Introduced |
|---|---|---|
| LCRV | (Radius of) Curvature | <2009 |
| LFAL | Crossfall | <2009 |
| LGRD | Gradient | <2009 |
| LV3 | 3m moving average LPV (left / nearside) | <2009 |
| LL03 | 3m enhanced LPV (nearside) | 2009 |
| LV10 | 10m moving average LPV (nearside) | <2009 |
| LL10 | 10m enhanced LPV (nearside) | 2009 |
| LLBI | Bump intensity (nearside) | 2009 |
| LR03 | 3m enhanced LPV (offside) | 2009 |
| LR10 | 10m enhanced LPV (offside) | 2009 |
| LRBI | Bump intensity (offside) | 2009 |
| LLRT | Nearside wheel path rut depth | <2009 |
| LLRD | Nearside rut depth from cleaned profile | 2009 |
| LRRT | Offside wheel path rut depth | <2009 |
| LRRD | Offside rut depth from cleaned profile | 2009 |
| LTAD | Absolute deviation of 1$^{st}$ derivative of transverse profile | 2009 |
| LTRV | Transverse variance | 2009 |
| LEDR | Edge roughness | 2009 |
| LES1 | Road edge step L1 (between 20 and 50mm step down) | 2009 |
| LES2 | Road edge step L2 (greater than 50mm step down) | 2009 |
| LEDC | Edge coverage | 2009 |
| LLTX | Nearside Wheel Path Average Texture depth (SMTD) | <2009 |

| UKPMS code | SCANNER survey parameter | Introduced |
|---|---|---|
| LLTD | Nearside Wheel Path Average Texture depth (MPD) | <2009 |
| LLTM | Nearside Wheel Path Mean RMST Texture depth | 2009 |
| LLTV | Nearside Wheel Path RMST Variance | 2009 |
| LCTM | Centre Mean RMST Texture depth | 2009 |
| LCTV | Centre RMST Variance | 2009 |
| LRTM | Offside Wheel Path Mean RMST Texture depth | 2009 |
| LRTV | Offside Wheel Path RMST Variance | 2009 |
| LT05 | Overall Texture Variability – RMST 5$^{th}$ Percentile Value | 2009 |
| LT95 | Overall Texture Variability – RMST 95$^{th}$ Percentile Value | 2009 |
| LTVV | Overall Texture Variability – RMST Variance | 2009 |
| LTRC | Cracking (whole carriageway) | <2009 |
| LWCL | Nearside Wheel Track Cracking Intensity | <2009 |
| LWCR | Offside Wheel Track Cracking Intensity | <2009 |
| LECR | Edge of carriageway cracking | <2009 |
| LOVD | Other Visible Defect | <2009 |
| LRCR | Transverse/reflection cracking | <2009 |
| LSUR | Surface Deterioration Parameter | <2009 |
| LSPD | Survey speed | <2009 |

The measurements used to calculate these parameters and the pavement features that they describe are given in the following sub-sections.

### *D.4.1 Road Roughness / Shape*

The longitudinal profile is the shape of the road in the direction of travel. SCANNER measures longitudinal profile in both the nearside and offside wheel paths. **Longitudinal profile variance (LPV)** is a measure of how much the road undulates. This is reported over 2 scales: 3m LPV and 10m LPV, where 3m LPV reports the undulation of the road due to features of less than 3m in length and 10m LPV reports undulation due to features of less than 10m in length. There is an **enhanced** version of these parameters which was developed to reduce the influence of road geometry on the reported roughness.

In addition to the general ride quality measures provided by LPV, SCANNER also reports the **Bump Measure** in the two wheelpaths, which indicates the presence of short features that cause discomfort to the users through bumping or jolting.

SCANNER also measures the **geometry**, reported as the gradient, the cross-fall and the radius of curvature of the road.

### D.4.2 Rutting and Transverse unevenness

The transverse profile is the shape of the road perpendicular to the direction of travel. The SCANNER measurements of the transverse profile are analysed to produce the parameters of rutting, transverse profile unevenness and edge condition.

**Rut depth** determined from SCANNER surveys corresponds to a measurement made with a 2m straight edge and wedge and average rut depths in the left (or nearside) and right (or offside) wheel paths are provided. There is an **enhanced** version of rutting which was developed to reduce the influence of the road edge on the reported rutting.

SCANNER also reports **Transverse Profile Unevenness**, which can be used to quantify how much the slope of the transverse profile changes from point to point across the carriageway. **Transverse Variance** is a measure of the difference in the roughness (transversally) between the two halves of the measurement width.

### D.4.3 Texture

Texture can be separated into two groups – single line and multiple line texture.

The **SMTD** and **MPD** parameters are calculated from texture, measured in a single line in the nearside wheelpath, and can be used to provide an indication of the high-speed skidding resistance.

Multiple line texture measurements from between 3 and 40 lines across the carriageway width, including the nearside and offside wheel paths, and the line midway between them, are used to calculate nine **RMST** parameters:

- The variation of texture in the nearside wheel path (Mean RMST and Variance)
- The variation of texture in the centre of the road (Mean RMST and Variance)
- The variation of texture in the offside wheel path (Mean RMST and Variance)
- Overall Texture Variability - RMST 5th Percentile Value, 95th Percentile Value and Variance.

### D.4.4 Surface Deterioration

SCANNER measures cracking on the surface of the pavement, which is reported as the location of each crack identified in the form of a crack map. The cracks are analysed to produce the three derived SCANNER cracking parameters:

- **Whole carriageway cracking**, obtained by overlaying the crack map with a grid covering the whole survey width, and summing up the areas of the grid squares containing cracks.
- **Wheel track cracking** intensity is reported over the two tracks, each of width 0.8m, centred on the wheel paths.
- **Transverse/reflective cracking** is a measure that attempts to indicate if the cracking is mainly transverse. Cracking that occupies a short length along the road but a large width across the road results in higher values of this parameter.
- **Surface Deterioration** is a measure that attempts to indicate if the cracking is short and “spread out”, as might be the case if the defects look like fretting that has begun the develop into crack-like features.

### D.4.5 Edge Deterioration

SCANNER uses the measured transverse profile to estimate the extent and severity of the deterioration of the road edge, which is reported as three parameters:

**Edge Roughness** reports the roughness within a half metre wide strip adjacent to the road edge.

**Edge steps** (L1 and L2) assess the height of the stepping present within the transverse profile adjacent to the identified road edge with LS1 being the percentage of reporting length with small step down at the road edge (20 to 50mm) and LS2 the percentage with large step down (greater than 50mm).

The Edge Coverage indicates the percentage of the reporting length where the profiles have been measured over the edge of the road. Where the value is low less confidence should be placed, in particular, on the measure of edge stepping.

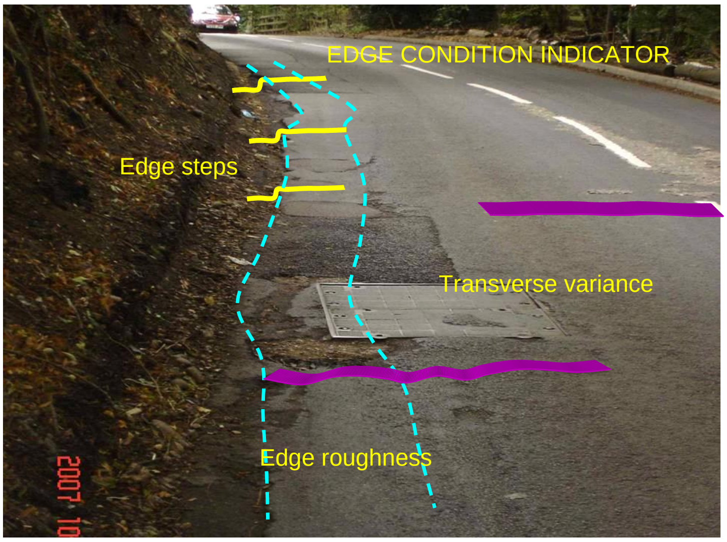



**Illustration of the components of the edge condition indicator**

### *D.4.6 The SCANNER RCI*

The primary aim of the SCANNER RCI, since 2005/06, has been to process SCANNER data to produce performance indicators. Currently the performance indicators produced using the RCI are the data topics 130-01 and 130-02 for the England Single Data List, the SRMCS PI for Scotland, THS/011 and THS/012 for Wales and a performance indicator for DRD Northern Ireland. Note that SCANNER data is also used for calculating depreciation but this is via a different calculation (referred to as the CCI).

The RCI is based on the following SCANNER parameters:

- LLRT: Nearside wheel path rut depth
- LRRT: Offside wheel path rut depth
- LLTX: Nearside Wheel Path Average Texture depth (SMTD)
- LTRC: Cracking (whole carriageway)
- LV10: 10m moving average LPV (nearside)
- LV3: 3m moving average LPV (left / nearside)

These parameters are weighted using a straight line between upper and lower thresholds which vary by road classification (and for texture, the thresholds also vary by rural/urban categorisation) and are combined to give an overall score for each subsection. Each subsection is then categorised as Red, Amber or Green based on this score.

# Appendix E Presentation to SCANNER Development Group, September 2016: Tasks 1 and 2

Development of SCANNER and UKPMS

**TASK 1: CONSISTENCY OF SCANNER DATA**



Page ▪ 1

## Task 1: Consistency

| Task | Activity | J | F | M | A | M | J | J | A | S | O | N | D | J | F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **Consistency of SCANNER data** | | | | | | | | | | | | | | |
| 1.1 | Review cracking accreditation process and tests and determine strengths and weaknesses | | | | 1a[3] | | | | | | | | | | |
| 1.2 | Develop cracking consistency assessment process | | | | | | | | | | | | | | |
| 1.3 | Trial using previous data and consult with stakeholders on the results | | | | | | | | | | | | | | |
| 1.5 | Review transverse profile data from network survey and accreditation tests, quantify issues | | | | 1a | | | | | | | | | | |
| 1.6 | With survey industry, propose how existing systems could deliver better data | | | | | | | | | | | | | | |
| 1.7 | Secure enhanced data and use to develop improved processing algorithms | | | | | | | | | | | | | | |
| 1.8 | Produce updated specification/algorithms | | | | | | | | | | | | 1b(i) | | |
| 1.9 | Provide advice to SCANNER auditor | | | | | | | | | | | | | | 1b(ii) |



Page ▪ 2

**TASK 1: CONSISTENCY OF CRACKING**



Page ▪ 3

## Task 1: Consistency - Cracking

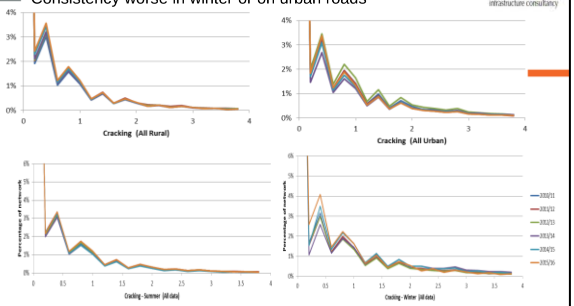
- Inconsistencies seen across the SCANNER fleet
- Can have a large effect on network reporting
- Aspects of survey affecting performance:
  - Urban roads show more variation than other road classifications
  - Less consistency when survey carried out in winter
- No evidence to support need for change to RCI thresholds
- Accreditation tests appropriate but need additional tests of absolute cracking values and fleet consistency.



Page ▪ 4

## Task 1: Consistency - Cracking

Consistency worse in winter or on urban roads

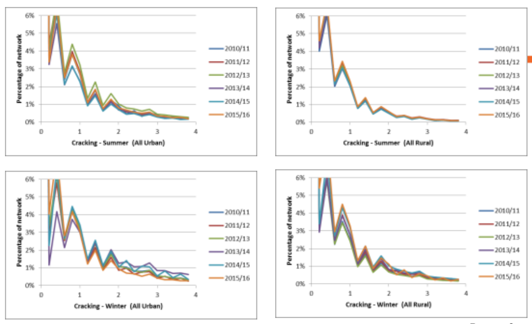



Investigate whether inconsistency is due to just urban lengths or just winter



Page ▪ 5

## Task 1: Consistency - Cracking






Page ▪ 6

## Task 1: Consistency – Cracking – Winter shutdown

- Consulted a selection of LAs for their thoughts on winter shutdown and whether cracking should always be included in the RCI:
  - Bristol, Carmarthenshire, Cornwall, Cumbria, Essex, Leicester, South Lanarkshire, Worcestershire.


Page • 7

## Task 1: Consistency – Cracking – Winter shutdown

- Winter shutdown:
  - Asked for their thoughts on a winter shutdown approach to improve the consistency of cracking data
  - Asked about effect on when data usually received
  - Asked what increase in cost would be found acceptable.
- Calculating RCI without cracking:
  - Asked for thoughts on excluding cracking from RCI, where measurements not reliable
  - Asked for thoughts on including an estimate of cracking from previous years


Page • 8

## Task 1: Consistency – Cracking – Thoughts on winter shutdown approach

- Four of the authorities consulted only had summer surveys, so a winter shutdown would not affect them/improve their data.
- Of the six authorities that gave an opinion (2 who have summer surveys) most felt that a winter shutdown would be a good idea, if it improved cracking consistency. One felt that this was a sticking plaster approach and we should just return to using CVI.
- Ensuring data was collected in the summer would improve the timelines, in terms of receiving the data in order to generate their programme of works
- All would prefer to see no cost increase but accepted costs ranged from 1-2% up to 5%.


Page • 9

## Task 1: Consistency – Cracking – Thoughts on cracking inclusion in RCI

- General unease at the suggestion of excluding cracking from RCI calculation when unreliable.
- Most (6) would prefer to avoid this and would prefer for fundamental issue with cracking to be resolved: worries about inconsistency in the RCI that this would cause.
- Including an estimate of cracking from previous year's data was also met with apprehension. Sticking plasters were mentioned again.


Page • 10

## Task 1: Cracking Accreditation

- Assessment of devices against reference still appropriate
- Identified need to enhance the test for repeatability of an individual device
- Identified the need to assess the consistency of the fleet


Page • 11

## Task 1: Cracking Accreditation – Repeatability

- Current approach scales the two runs by their average and standard deviation
- Appropriate when comparing reference with machine data but really a device should be expected to provide the same values for cracking on repeat runs.
- Considered a number of approaches:
  - SCRIM method
  - Chris Britton method used to calculate SCANNER consistency
  - Alternative but similar consistency method

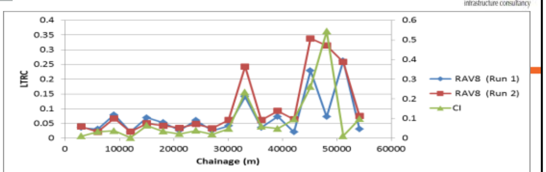

Page • 12

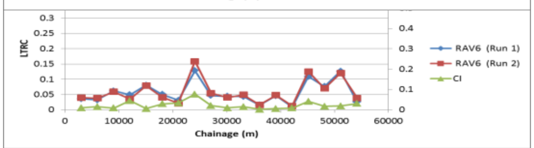
## Task 1: Cracking Accreditation – Repeatability

- Confidence interval method:
  - Assume that the errors from a device have a normal distribution
  - Possible to calculate a 95% Confidence Interval around the best estimate of the population average
  - Standard calculation used for this value
- Advantages over simple difference calculation:
  - Can be used on data from more than two repeat runs
  - Does not require one run to act as the "reference".


Page • 13

## Task 1: Cracking Accreditation – Repeatability – applying the confidence interval method

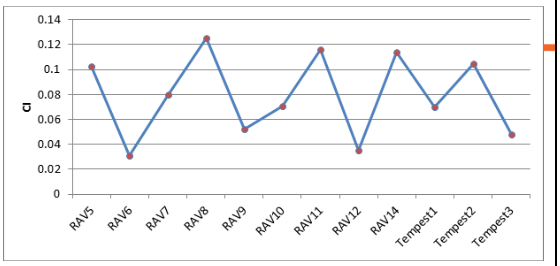




Page • 14

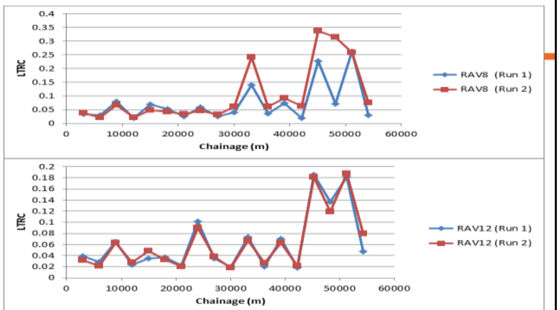
## Task 1: Cracking Accreditation – Repeatability – average confidence intervals





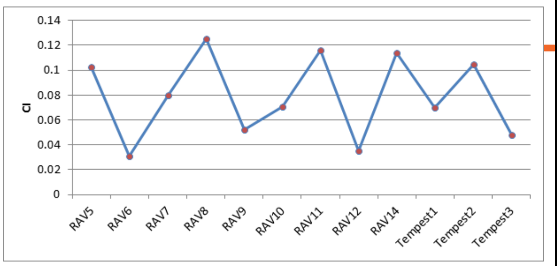


## Task 1: Cracking Accreditation – Repeatability – applying the CI method






## Task 1: Cracking Accreditation – Repeatability – average confidence intervals







## Task 1: Cracking Accreditation – Fleet consistency

- Considered a number of different statistical parameters when looking at fleet consistency including:
  - Mean
  - Standard error
  - Standard deviation
  - Skewness
  - Range
  - Bias

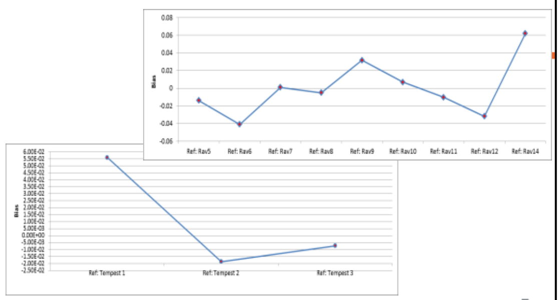



## Task 1: Cracking Accreditation – Fleet consistency

- Settled on the average bias of each device from the fleet:
  - Calculate the biases of a device from each of the other devices in the fleet
  - Calculate the average of the absolute value of these biases
  - Calculate the average of the biases – the sign of this is the sign of the resulting bias.

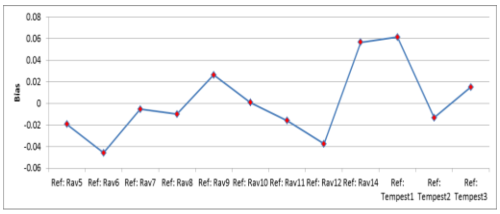



## Task 1: Cracking Accreditation – Fleet Consistency






## Task 1: Cracking Accreditation – Fleet Consistency







## Task 1: Cracking Accreditation – Fleet consistency

- Approach to applying the fleet consistency test:
  - Ideally would like to apply a maximum (absolute) bias for an individual device in the fleet to be 0.01…
  - Might consider 0.04 initially, tightening to 0.02 over an agreed period of time.
  - Would not be used to fail individual machines – contractors will choose how to improve the consistency of their fleet.

# TASK 1: CONSISTENCY OF RUTTING




## Task 1: Consistency - Rutting

- This task to confirm/investigate the actual consistency of rutting and to consider how it could be improved
- To investigate whether enhancements in technology could deliver improvements in SCANNER rutting
- Whether experience from TRACS and HARRIS2 could be used to improve SCANNER rutting.

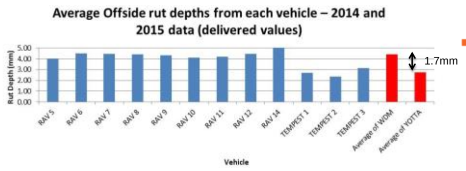



## Task 1: Consistency - Rutting – Summary

- Inconsistency in rutting seen in LAs year on year
- Rutting has large effect on RCI
  - But effect of inconsistency for routine surveys still needs quantifying
- Inconsistency is greater in the nearside
- Inconsistency is present across the fleet (bias between fleets found on accreditation site)



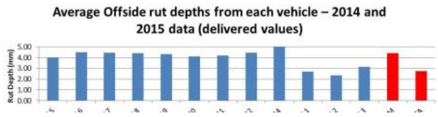

## Task 1: Rutting – Fleet Consistency

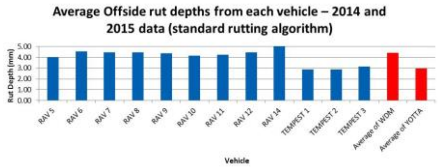






## Task 1: Rutting – Fleet Consistency







## Task 1: Potential improvements to the rutting measurement

- The possible areas of improvement to the rut algorithm could be:
  - Cleaned rutting
  - Use of high resolution profile
    - Removal of road markings
    - Removal of kerbs
    - Removal of other edge features (verges etc.)
    - Increased width of profile
    - Minimum valid width of profile

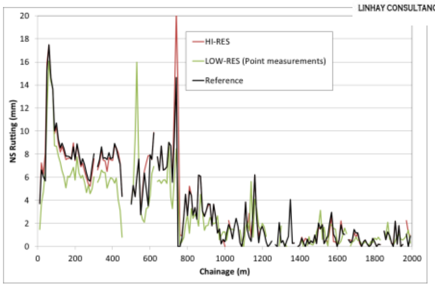



## Task 1: Accuracy of Rut depths

- Rutting calculated using high resolution transverse profile data more consistent than that from low resolution.
- Suggestion that a less consistent but more accurate rut parameter would actually be better.
- Investigation of rut depth accuracy on SRR2
  - Not within scope to collect true reference (i.e. straight edge/wedge) measurements
  - Used a manual assessment of forward facing images and transverse profile to verify results.

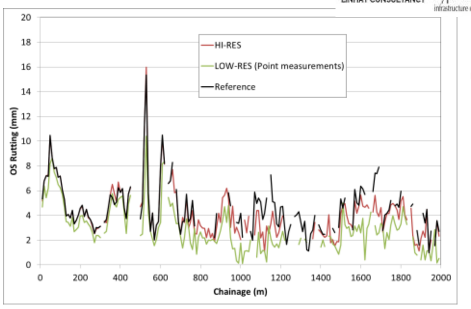



## Accuracy of Rut depths - nearside




- Both results calculated using TRACS rutting algorithm
- High resolution gives a better match to reference

## Accuracy of Rut depths - offside







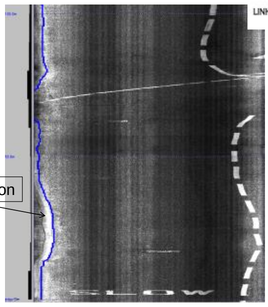
## Task 1: Cleaned rutting investigation

- Previous analysis showed that cleaned rutting actually appeared to be less consistent than standard rutting and could sometimes give some unexpected values.
- Investigated this
  - Found that it was due to the edge identification




## Task 1: Edge detection

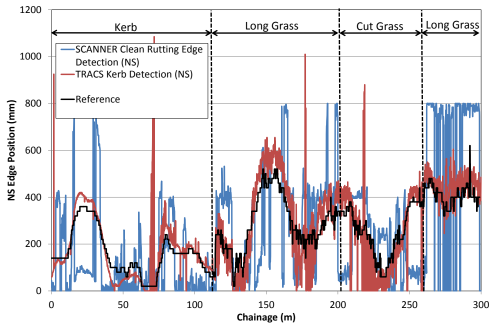
Edge position



Manual analysis of downward and forward facing images to obtain reference edge position data




## Task 1: Edge detection







## Task 1: Edge detection

- The TRACS edge position (red) is pretty good and matches the manual assessment well.
- The cleaned rutting edge position is quite bonkers in places.
- Suprising: TRACS is very basic and just looks for unexpected changes in height, whereas the cleaned rutting algorithm is much more sophisticated.
- Lengths where edge incorrectly detected by cleaned rutting are where issues with the consistency of the cleaned rutting are also seen.
- Key to obtaining an accurate and consistent measure of the rut depths on a road is for the **edge to be detected well**.

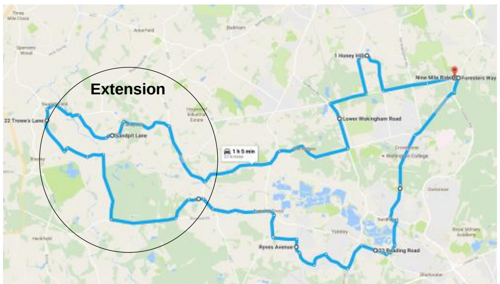



## Task 1: Further investigation with high resolution data

- Detailed analysis can only be carried out with raw data
- No raw data received from contractors to date
- Planning to survey an extension to SCANNER Road Route 2 with HARRIS2 which includes:
  - U roads
  - Flat grass verges
  - Hedges
  - Raised verges
- Perform manual analysis of the data to create a reference dataset




## Task 1: SCANNER Road Route Extension







## Task 1: SCANNER Road Route Extension

## Task 1: Where to next

- **Cracking work will consider**:
  - How to apply repeatability test and what threshold to apply to the CI
  - How to apply fleet consistency test and what threshold to apply to the average bias
- **Rutting work:**
  - Survey Extended SRR2
  - Investigate edge detection on roads with “non-trunk road” edges e.g. verges, hedges using this data




Development of SCANNER and UKPMS

## TASK 2: SCANNER PARAMETERS

Deliverable 2b: Progress Report and Presentation




## Development of SCANNER and UKPMS Task 2: SCANNER Condition Parameters

- Objective:
  - To determine how the parameters provided by SCANNER could be optimised to meet LA needs (and hence provide better value)
  - To develop “quick wins” where feasible
  - To identify the requirements for longer term development of the parameter set
- Issue:
  - There is concern that SCANNER does not report all of the defects that authorities regard as important to include in a condition survey
    - SCANNER delivers >20 parameters. Only a few are used in the RCI
    - Few LAs make use of the enhanced parameters provided in the 2007 research.
    - The survey does not report some parameters considered important.
- Task:
  - Led by Emma Benbow, who leads much of our research into the development of TRACS




## Task 2: SCANNER Condition Parameters

| Task | Activity | J | F | M | A | M | J | J | A | S | O | N | D | J | F |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2** | **The SCANNER Condition Parameters** | | | | | | | | | | | | | | |
| 2.1 | Consult with LHAs on SCANNER parameter set | | | | 2a | | | | | | | | | | |
| 2.2 | Review SCANNER raw data and propose potential quick win enhancements | | | | | | | | | | | | | | |
| 2.3 | Present summary quick win proposals to SDG and agree development area(s) | | | | | | | | | | | | | | |
| 2.4 | Undertake development of processes/algorithms for quick wins | | | | | | | | 2b | | | | | | |
| 2.5 | Produce updated specifcation/algorithms | | | | | | | | | | | | | | 2c |
| 2.6 | Provide advice to SCANNER auditor | | | | | | | | | | | | | | |
| 2.7 | Review current and emerging technology, determine potential for development/improvement | | | | | | | | | | | | | | |
| 2.8 | Determine long term development potential and produce development programme | | | | | | | | | | | | 2d | | |

## Task 2: SCANNER Condition Parameters: Approach

- Consult with LAs/PMS providers to better understand current use of SCANNER parameters: Questionnaire, in parallel with Task 3
- Categorise the SCANNER parameters
  - ‘valuable/essential’; ‘moderate use’; ‘worth developing/adding’; ‘little use/unreliable’; ‘important but not provided’ etc.
  - Shortlist areas for development (requirement)
- Review current/new technologies, to determine improvement categories such as:
  - ‘quick-win achievable’; ‘deliverable (further-research)’; ‘remove’ etc.
  - Hence shortlisting areas for development (technology)
- Agree the quick win(s) and develop these (within scope of project).
- Where development is outside scope, produce detailed programme of the work required.




## Task 2: Parameters: Consultation

- Questionnaire sent to 35 recipients:
  - 29 from England, of which 8 metropolitan or London borough authorities
  - 3 from Scotland, 2 from Wales,1 from Ireland
- 16 responses, but only 15 including responses for parameters, of which
  - 11 from England, 3 from Scotland, 1 from Wales, 1 from Northern Ireland
- Responses from England:
  - 7 from counties, 2 from unitary authorities, 1 from a PFI, 1 from London borough




## Task 2: Initial analysis of consultation

- Summary:
- Most important parameters seemed to be those used in RCI calculation: Rutting, Cracking, 3m LPV, Texture, 10m LPV.
- Most didn’t seem to know about or understand the enhanced parameters – very little use being made.




## Task 2: Recommendations from consultation

- Quick Wins:
- Phase out use of LPV and replace with eLPV;
- Drop cleaned rutting and continue with “standard” rutting;
- Replace normal rut with “better” measure (Task 1);
- Drop all cracking parameters except Whole Carriageway Cracking;
- Drop Other Visible Defects parameter;
- Keep Edge parameters;
- Investigate suitability of Bump Measure to provide a pothole indicator

## Task 2: Recommendations from consultation

- Long term development:
- Multiple line texture – drop and develop fretting?
- Investigate the capability of SCANNER to provide a better bump/pothole measure.
- Need of education for users.




## Task 2: Investigation of eLPV

- Determine whether eLPV is more or less consistent than LPV in the SCANNER environment:
  - Effect of geometry on the measures
  - Consistency from Accreditation tests
  - QA Audit reports
- Investigate the step change that would be seen if a move to using eLPV was implemented.




## Task 2: Effect of geometry on LPV

- TRACS: LPV effected by geometry, particularly 10m LPV
  - Unexpectedly high values on bends, and where high crossfall present
- Used data from Devon to investigate the effect on local roads.
- Categorisation of lengths as Green/Amber/Red
  - Used RCI thresholds for LPV
  - Calculated thresholds for eLPV using TRACS thresholds.




## Task 2: eLPV compared with LPV

Table shows a length by length comparison of the category that the length would be placed in by either LPV or eLPV

| | | LPV | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Class A | | | Class B | | | Class C | | | All roads | | |
| | | G | A | R | G | A | R | G | A | R | G | A | R |
| 3m eLPV | G | 22.6 | 0.3 | 0.0 | 14.5 | 0.3 | 0.0 | 47.3 | 3.4 | 0.1 | 84.4 | 4.0 | 0.1 |
| | A | 0.2 | 0.7 | 0.1 | 0.2 | 0.9 | 0.0 | 0.9 | 5.7 | 0.6 | 1.3 | 7.3 | 0.7 |
| | R | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.2 | 0.0 | 0.3 | 1.4 | 0.0 | 0.4 | 1.8 |
| 10m eLPV | G | 19.7 | 2.5 | 0.6 | 12.1 | 2.1 | 0.5 | 31.9 | 10.7 | 3.1 | 63.8 | 15.3 | 4.2 |
| | A | 0.2 | 0.6 | 0.3 | 0.2 | 0.6 | 0.3 | 1.5 | 6.1 | 3.8 | 1.9 | 7.3 | 4.4 |
| | R | 0.0 | 0.1 | 0.2 | 0.0 | 0.1 | 0.2 | 0.0 | 0.5 | 2.2 | 0.0 | 0.6 | 2.6 |




## Task 2: eLPV compared with LPV

Can see that more lengths are placed in higher categories by LPV when consider all roads, particularly for 10m LPV

| | | LPV | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Class A | | | Class B | | | Class C | | | All roads | | |
| | | G | A | R | G | A | R | G | A | R | G | A | R |
| 3m eLPV | G | 22.6 | 0.3 | 0.0 | 14.5 | 0.3 | 0.0 | 47.3 | 3.4 | 0.1 | 84.4 | 4.0 | 0.1 |
| | A | 0.2 | 0.7 | 0.1 | 0.2 | 0.9 | 0.0 | 0.9 | 5.7 | 0.6 | 1.3 | 7.3 | 0.7 |
| | R | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.2 | 0.0 | 0.3 | 1.4 | 0.0 | 0.4 | 1.8 |
| 10m eLPV | G | 19.7 | 2.5 | 0.6 | 12.1 | 2.1 | 0.5 | 31.9 | 10.7 | 3.1 | 63.8 | 15.3 | 4.2 |
| | A | 0.2 | 0.6 | 0.3 | 0.2 | 0.6 | 0.3 | 1.5 | 6.1 | 3.8 | 1.9 | 7.3 | 4.4 |
| | R | 0.0 | 0.1 | 0.2 | 0.0 | 0.1 | 0.2 | 0.0 | 0.5 | 2.2 | 0.0 | 0.6 | 2.6 |




## Task 2: eLPV compared with LPV

Not so marked for A and B roads

| | | LPV | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Class A | | | Class B | | | Class C | | | All roads | | |
| | | G | A | R | G | A | R | G | A | R | G | A | R |
| 3m eLPV | G | 22.6 | 0.3 | 0.0 | 14.5 | 0.3 | 0.0 | 47.3 | 3.4 | 0.1 | 84.4 | 4.0 | 0.1 |
| | A | 0.2 | 0.7 | 0.1 | 0.2 | 0.9 | 0.0 | 0.9 | 5.7 | 0.6 | 1.3 | 7.3 | 0.7 |
| | R | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.2 | 0.0 | 0.3 | 1.4 | 0.0 | 0.4 | 1.8 |
| 10m eLPV | G | 19.7 | 2.5 | 0.6 | 12.1 | 2.1 | 0.5 | 31.9 | 10.7 | 3.1 | 63.8 | 15.3 | 4.2 |
| | A | 0.2 | 0.6 | 0.3 | 0.2 | 0.6 | 0.3 | 1.5 | 6.1 | 3.8 | 1.9 | 7.3 | 4.4 |
| | R | 0.0 | 0.1 | 0.2 | 0.0 | 0.1 | 0.2 | 0.0 | 0.5 | 2.2 | 0.0 | 0.6 | 2.6 |




## Task 2: eLPV compared with LPV

Most noticeable on C roads, where we would expect extremes of geometry

| | | LPV | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Class A | | | Class B | | | Class C | | | All roads | | |
| | | G | A | R | G | A | R | G | A | R | G | A | R |
| 3m eLPV | G | 22.6 | 0.3 | 0.0 | 14.5 | 0.3 | 0.0 | 47.3 | 3.4 | 0.1 | 84.4 | 4.0 | 0.1 |
| | A | 0.2 | 0.7 | 0.1 | 0.2 | 0.9 | 0.0 | 0.9 | 5.7 | 0.6 | 1.3 | 7.3 | 0.7 |
| | R | 0.0 | 0.1 | 0.2 | 0.0 | 0.0 | 0.2 | 0.0 | 0.3 | 1.4 | 0.0 | 0.4 | 1.8 |
| 10m eLPV | G | 19.7 | 2.5 | 0.6 | 12.1 | 2.1 | 0.5 | 31.9 | 10.7 | 3.1 | 63.8 | 15.3 | 4.2 |
| | A | 0.2 | 0.6 | 0.3 | 0.2 | 0.6 | 0.3 | 1.5 | 6.1 | 3.8 | 1.9 | 7.3 | 4.4 |
| | R | 0.0 | 0.1 | 0.2 | 0.0 | 0.1 | 0.2 | 0.0 | 0.5 | 2.2 | 0.0 | 0.6 | 2.6 |




## Task 2: Geometry classification

- Need to look at whether it's geometry causing the high levels of LPV
- 5 classes developed for curvature, crossfall, gradient
- Curvature:
  - Straight (category 1): greater than 1000m
  - Moderately curved (category 2): between 500m and 1000m
  - Fairly curved (category 3): between 100m and 500m
  - Curved (category 4): between 50m and 100m
  - Extremely curved (category 5): less than 50m.

## Task 2: Geometry classification

- Crossfall:
  - Adequate crossfall (category 1): less than 2%
  - Slightly higher than adequate crossfall (category 2): between 2% and 3%;
  - Moderately high crossfall (category 3): between 3% and 7%
  - High crossfall (category 4): between 7% and 10%
  - Very high crossfall (category 5): greater than 10% .




## Task 2: Geometry classification

- Gradient:
  - Flat (category 1): less than 2%
  - Low gradient (category 2): between 2% and 5%
  - Moderately steep (category 3): between 5% and 10%
  - Steep (category 4): between 10% and 15%
  - Very steep (category 5): greater than 10%




## Task 2: 3m eLPV and LPV, split by road class and geometry classification

On A and B roads: eLPV and LPV match fairly well on lengths without much geometry (categories 1-3).

LPV always higher for lengths in higher categories

| Class | Geometry category | | Curvature G | Curvature A | Curvature R | Crossfall G | Crossfall A | Crossfall R | Gradient G | Gradient A | Gradient R |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Class A | 1 | 3m LPV | 48.7 | 1.62 | 0.46 | 30 | 1.46 | 0.39 | 39.1 | 1.56 | 0.45 |
| | 2 | | 16.3 | 0.78 | 0.2 | 25.6 | 0.97 | 0.27 | 33.9 | 1.46 | 0.42 |
| | 3 | | 25.9 | 1.61 | 0.5 | 36.1 | 1.86 | 0.57 | 19.1 | 1.2 | 0.35 |
| | 4 | | 2.69 | 0.29 | 0.07 | 2.44 | 0.14 | 0.06 | 2.05 | 0.22 | 0.06 |
| | 5 | | 0.72 | 0.17 | 0.06 | 0.16 | 0.02 | 0.01 | 0.13 | 0.01 | 0.01 |
| | 1 | 3m eLPV | 48.7 | 1.51 | 0.47 | 30.2 | 1.24 | 0.38 | 39.2 | 1.42 | 0.45 |
| | 2 | | 16.4 | 0.67 | 0.23 | 25.7 | 0.87 | 0.28 | 34.1 | 1.27 | 0.43 |
| | 3 | | 26.1 | 1.44 | 0.48 | 36.2 | 1.71 | 0.56 | 19.3 | 1.04 | 0.33 |
| | 4 | | 2.77 | 0.21 | 0.07 | 2.48 | 0.11 | 0.04 | 2.07 | 0.2 | 0.06 |
| | 5 | | 0.81 | 0.11 | 0.03 | 0.17 | 0.01 | 0 | 0.13 | 0.01 | 0 |
| Class B | 1 | 3m LPV | 39.2 | 2.43 | 0.45 | 34.1 | 2.73 | 0.58 | 35.4 | 2.77 | 0.61 |
| | 2 | | 15.8 | 1.28 | 0.26 | 20.8 | 1.5 | 0.32 | 31 | 2.5 | 0.49 |
| | 3 | | 31.1 | 2.88 | 0.6 | 34 | 2.97 | 0.62 | 20.8 | 1.83 | 0.4 |
| | 4 | | 3.58 | 0.57 | 0.14 | 1.91 | 0.31 | 0.06 | 3.25 | 0.37 | 0.08 |
| | 5 | | 1.11 | 0.37 | 0.13 | 0.09 | 0.01 | 0 | 0.39 | 0.06 | 0.01 |
| | 1 | 3m eLPV | 39.3 | 2.33 | 0.45 | 34.3 | 2.58 | 0.53 | 35.6 | 2.61 | 0.59 |
| | 2 | | 15.9 | 1.23 | 0.27 | 20.9 | 1.4 | 0.32 | 31.3 | 2.26 | 0.44 |
| | 3 | | 31.4 | 2.6 | 0.58 | 34.4 | 2.65 | 0.6 | 21.1 | 1.61 | 0.39 |
| | 4 | | 3.67 | 0.49 | 0.13 | 1.95 | 0.27 | 0.07 | 3.26 | 0.36 | 0.08 |
| | 5 | | 1.28 | 0.26 | 0.07 | 0.09 | 0.01 | 0 | 0.39 | 0.07 | 0.01 |




## Task 2: 10m eLPV and LPV, split by road class and geometry classification

Similarly for 10m LPV but even more noticeable

| Class | Geometry category | | Curvature G | Curvature A | Curvature R | Crossfall G | Crossfall A | Crossfall R | Gradient G | Gradient A | Gradient R |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Class A | 1 | 10m LPV | 44.7 | 4.9 | 1.08 | 26 | 4.37 | 1.43 | 35.4 | 4.47 | 1.24 |
| | 2 | | 14.5 | 2.23 | 0.63 | 23.1 | 2.86 | 0.85 | 29.6 | 4.7 | 1.45 |
| | 3 | | 21.2 | 4.94 | 1.87 | 31.2 | 5.46 | 1.82 | 15.8 | 3.46 | 1.42 |
| | 4 | | 1.7 | 0.85 | 0.5 | 1.89 | 0.46 | 0.3 | 1.47 | 0.54 | 0.32 |
| | 5 | | 0.29 | 0.28 | 0.38 | 0.07 | 0.06 | 0.07 | 0.08 | 0.04 | 0.03 |
| | 1 | 10m eLPV | 48.8 | 1.6 | 0.33 | 30 | 1.53 | 0.28 | 39.2 | 1.59 | 0.34 |
| | 2 | | 16.4 | 0.78 | 0.14 | 25.7 | 0.94 | 0.2 | 34 | 1.49 | 0.3 |
| | 3 | | 26 | 1.65 | 0.38 | 36.2 | 1.82 | 0.44 | 19.2 | 1.19 | 0.27 |
| | 4 | | 2.7 | 0.28 | 0.07 | 2.45 | 0.16 | 0.03 | 2.09 | 0.18 | 0.05 |
| | 5 | | 0.75 | 0.16 | 0.05 | 0.17 | 0.02 | 0 | 0.13 | 0.01 | 0 |
| Class B | 1 | 10m LPV | 34.7 | 6 | 1.37 | 28.8 | 6.49 | 2.14 | 31.3 | 5.75 | 1.75 |
| | 2 | | 13.7 | 2.84 | 0.82 | 17.9 | 3.57 | 1.15 | 26.1 | 5.88 | 2.01 |
| | 3 | | 25.3 | 6.8 | 2.51 | 28.4 | 6.71 | 2.47 | 16.6 | 4.58 | 1.88 |
| | 4 | | 2.26 | 1.23 | 0.8 | 1.36 | 0.56 | 0.36 | 2.28 | 0.97 | 0.45 |
| | 5 | | 0.49 | 0.47 | 0.65 | 0.04 | 0.03 | 0.04 | 0.24 | 0.17 | 0.06 |
| | 1 | 10m eLPV | 39.5 | 2.25 | 0.36 | 34.4 | 2.53 | 0.47 | 35.6 | 2.63 | 0.52 |
| | 2 | | 16 | 1.16 | 0.21 | 20.9 | 1.46 | 0.28 | 31.2 | 2.34 | 0.45 |
| | 3 | | 31.3 | 2.73 | 0.57 | 34.3 | 2.75 | 0.61 | 21 | 1.71 | 0.36 |
| | 4 | | 3.6 | 0.55 | 0.14 | 1.93 | 0.28 | 0.08 | 3.32 | 0.31 | 0.08 |
| | 5 | | 1.13 | 0.34 | 0.15 | 0.09 | 0.01 | 0 | 0.41 | 0.05 | 0.01 |




## Task 2: eLPV and LPV, split by geometry classification for C roads

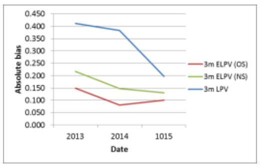
More of the network falls into high geometry categories for C roads, so difference between the two parameters is more noticeable

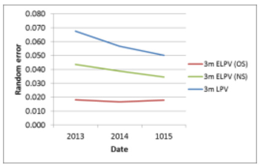
| Class | Geometry category | | Curvature G | Curvature A | Curvature R | Crossfall G | Crossfall A | Crossfall R | Gradient G | Gradient A | Gradient R |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Class C | 1 | 3m LPV | 31 | 4.19 | 0.7 | 40.4 | 7.82 | 1.63 | 27.5 | 4.08 | 0.9 |
| | 2 | | 15 | 2.58 | 0.48 | 16.8 | 2.95 | 0.63 | 27.5 | 4.58 | 0.94 |
| | 3 | | 29 | 6.44 | 1.41 | 22.8 | 4.61 | 1.12 | 20 | 4.56 | 1.01 |
| | 4 | | 4.13 | 1.41 | 0.4 | 0.9 | 0.24 | 0.09 | 5.21 | 1.94 | 0.49 |
| | 5 | | 1.72 | 1.01 | 0.48 | 0.03 | 0.01 | 0.01 | 0.73 | 0.47 | 0.15 |
| | 1 | 3m eLPV | 32 | 3.26 | 0.63 | 42.6 | 6.03 | 1.29 | 28.5 | 3.27 | 0.76 |
| | 2 | | 15.6 | 2.03 | 0.42 | 17.5 | 2.32 | 0.51 | 28.8 | 3.47 | 0.74 |
| | 3 | | 30.7 | 5 | 1.12 | 24.1 | 3.58 | 0.86 | 21.4 | 3.4 | 0.75 |
| | 4 | | 4.59 | 1.06 | 0.29 | 0.98 | 0.19 | 0.06 | 5.71 | 1.57 | 0.36 |
| | 5 | | 2.16 | 0.78 | 0.27 | 0.04 | 0.01 | 0 | 0.8 | 0.43 | 0.12 |
| Class C | 1 | 10m LPV | 23.5 | 9.17 | 3.26 | 27.6 | 14.9 | 7.38 | 21.4 | 7.96 | 3.15 |
| | 2 | | 10.7 | 5.25 | 2.16 | 12 | 5.6 | 2.74 | 19 | 9.56 | 4.42 |
| | 3 | | 19 | 11.6 | 6.26 | 15.8 | 7.99 | 4.71 | 12.3 | 8.15 | 5.08 |
| | 4 | | 2.2 | 1.93 | 1.81 | 0.57 | 0.34 | 0.32 | 2.85 | 2.7 | 2.09 |
| | 5 | | 0.64 | 0.91 | 1.67 | 0.01 | 0.02 | 0.03 | 0.44 | 0.47 | 0.43 |
| | 1 | 10m eLPV | 29.5 | 5.47 | 0.97 | 38.1 | 9.63 | 2.16 | 26.3 | 5.14 | 1.09 |
| | 2 | | 14.2 | 3.26 | 0.65 | 15.9 | 3.61 | 0.81 | 25.9 | 5.86 | 1.26 |
| | 3 | | 27.3 | 7.7 | 1.87 | 21.6 | 5.42 | 1.45 | 18.8 | 5.41 | 1.37 |
| | 4 | | 3.91 | 1.52 | 0.51 | 0.85 | 0.28 | 0.11 | 4.89 | 2.1 | 0.65 |
| | 5 | | 1.67 | 1 | 0.55 | 0.03 | 0.01 | 0.01 | 0.73 | 0.44 | 0.18 |




## Task 2: Consistency of eLPV – Accreditation

- Consistency of each parameter calculated each year and results used in Audit reports:
  - Bias (average difference between device and reference)
  - Random error (error between repeat runs)

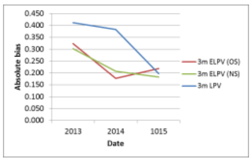


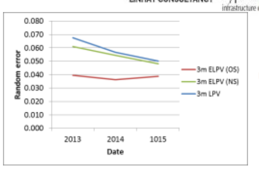



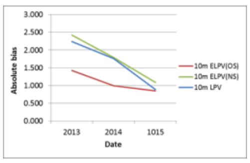

  - Not fair comparison: LPV larger values/range than eLPV
  - Have normalised eLPV in comparison with LPV

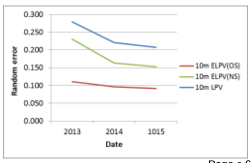



## Task 2: Consistency of eLPV – Accreditation







## Task 2: Consistency of eLPV – QA Audit Reports

- Have shown that eLPV is as consistent as LPV on the SCANNER road routes
- Investigate whether using on network would help with consistency there
- Looked at effect of using eLPV on Audit Reports for Devon and Herefordshire
  - Issues with consistency of LPV
  - Effect on RCI/Audit Indicator
  - Does using eLPV help?

## Task 2: Consistency of eLPV – QA Devon Audit Report (2013 and 2015 data)

LINHAY CONSULTANCY hyperion infrastructure consultancy

| Param. | Data | Mean | 99th percentile[1] | Distn. Corr. | Bias from Prev. | Expect Bias Range | Red[3] | Amber[3] | Green[3] |
|---|---|---|---|---|---|---|---|---|---|
| Nearside Rut | Curr. | 5.65 | 17.60 | 0.99 | 0.64 | -0.68 to 0.68 | 0.4% | 12.0% | 87.6% |
| | Prev. | 5.01 | 14.30 | | | | 0.1% | 6.7% | 93.2% |
| Offside Rut | Curr. | 5.44 | 17.30 | 0.97 | 1.16 | -0.52 to 0.52 | 0.4% | 11.5% | 88.1% |
| | Prev. | 4.29 | 13.70 | | | | 0.1% | 5.1% | 94.8% |
| Nearside 3m LPV | Curr. | 4.07 | 35.10 | 1.00 | 0.87 | -1.24 to 1.24 | 7.7% | 22.7% | 69.6% |
| | Prev. | 3.21 | 26.90 | | | | 4.6% | 17.0% | 78.4% |
| Nearside 10m LPV | Curr. | 32.91 | 611.29 | 1.00 | 4.57 | -2.26 to 2.26 | 22.9% | 25.1% | 52.0% |
| | Prev. | 28.35 | 438.60 | | | | 16.3% | 23.8% | 59.9% |
| Nearside SMTD | Curr. | 0.68 | 0.21 | 0.97 | -0.08 | -0.07 to 0.07 | 6.1% | 36.2% | 57.7% |
| | Prev. | 0.77 | 0.24 | | | | 3.5% | 28.3% | 68.2% |
| Cracking | Curr. | 0.34 | 3.80 | 0.99 | 0.16 | -0.17 to 0.17 | 5.1% | 29.7% | 65.2% |
| | Prev. | 0.18 | 2.20 | | | | 1.6% | 22.3% | 76.1% |
| RCI | Curr. | 49.09 | 180.32 | 0.99 | 16.70 | -12.29 to 12.29 | 13.9% | 37.3% | 48.7% |
| | Prev. | 32.39 | 135.00 | | | | 5.0% | 30.8% | 64.3% |

| Audit Indicator[4] | | | Expected Range of Indicator for Curr. year |
|---|---|---|---|
| | Curr. | 13.9% | |
| | Prev. | 5.0% | 3.5% to 6.5% |

[1] 1st Percentile for NS SMTD [2] Using RCI thresholds for urban B roads [4] Percentage of RCI values ≥ 100 (**Red** Category)
[3] Red = Max contribution to the RCI, Amber = Contributing to the RCI but not maximum, Green = Not contributing at all

**1 of 3 audit criteria met**

**Notes & Suggested Actions:**
Offside Rut, Nearside 10m LPV, Nearside SMTD and RCI are all outside expected bias from previous survey
The Audit Indicator is greater than expected for the current year

For further information please contact your current SCANNER survey provider.
If you then require additional information please email scanner@trl.co.uk.

Issues with the bias in 10m LPV, offside rutting, SMTD and the Audit Indicator




## Task 2: Consistency of eLPV – QA Devon Audit Report, using eLPV

LINHAY CONSULTANCY hyperion infrastructure consultancy

| Param. | Data | Mean | 99th percentile[1] | Distn. Corr. | Bias from Prev. | Expect Bias Range | Red[3] | Amber[3] | Green[3] |
|---|---|---|---|---|---|---|---|---|---|
| Nearside Rut | Curr. | 5.64 | 17.40 | 0.99 | 0.63 | -0.68 to 0.68 | 0.4% | 12.0% | 87.6% |
| | Prev. | 5.01 | 14.30 | | | | 0.1% | 6.7% | 93.2% |
| Offside Rut | Curr. | 5.44 | 17.20 | 0.97 | 1.15 | -0.52 to 0.52 | 0.4% | 11.4% | 88.2% |
| | Prev. | 4.29 | 13.70 | | | | 0.1% | 5.2% | 94.8% |
| Nearside 3m LPV | Curr. | 1.89 | 17.57 | 1.00 | 0.37 | -1.23 to 1.23 | 5.4% | 19.4% | 75.2% |
| | Prev. | 1.51 | 13.22 | | | | 3.4% | 14.4% | 82.3% |
| Nearside 10m LPV | Curr. | 9.42 | 87.99 | 1.00 | 1.98 | -2.2 to 2.20 | 10.0% | 21.9% | 68.0% |
| | Prev. | 7.44 | 67.88 | | | | 6.1% | 17.1% | 76.7% |
| Nearside SMTD | Curr. | 0.68 | 0.21 | 0.97 | -0.08 | -0.07 to 0.07 | 6.1% | 36.1% | 57.7% |
| | Prev. | 0.77 | 0.24 | | | | 3.5% | 28.4% | 68.1% |
| Cracking | Curr. | 0.34 | 3.80 | 0.99 | 0.16 | -0.17 to 0.17 | 5.1% | 29.7% | 65.2% |
| | Prev. | 0.18 | 2.20 | | | | 1.6% | 22.3% | 76.1% |
| RCI | Curr. | 40.45 | 168.11 | 0.99 | 15.32 | -12.29 to 12.29 | 9.2% | 33.6% | 57.2% |
| | Prev. | 25.13 | 124.80 | | | | 2.9% | 23.7% | 73.4% |

| Audit Indicator[4] | | | Expected Range of Indicator for Curr. year |
|---|---|---|---|
| | Curr. | 9.2% | |
| | Prev. | 2.9% | 1.6% to 4.1% |

[1] 1st Percentile for NS SMTD [2] Using RCI thresholds for urban B roads [4] Percentage of RCI values ≥ 100 (**Red** Category)
[3] Red = Max contribution to the RCI, Amber = Contributing to the RCI but not maximum, Green = Not contributing at all

**1 of 3 audit criteria met**

**Notes & Suggested Actions:**
Offside Rut, Nearside SMTD and RCI are all outside expected bias from previous survey
The Audit Indicator is greater than expected for the current year

For further information please contact your current SCANNER survey provider.
If you then require additional information please email scanner@trl.co.uk.

Calculated new ranges for the bias for eLPV using accreditation data.

Using eLPV sorts out issues with bias but not Audit indicator – still being influenced by rutting and texture.




## Task 2: Consistency of eLPV – QA Herefordshire Audit Report

LINHAY CONSULTANCY hyperion infrastructure consultancy

| Param. | Data | Mean | 99th percentile[1] | Distn. Corr. | Bias from Prev. | Expect Bias Range | Red[3] | Amber[3] | Green[3] |
|---|---|---|---|---|---|---|---|---|---|
| Nearside Rut | Curr. | 5.61 | 16.10 | 0.96 | -0.60 | -0.71 to 0.71 | 0.2% | 10.4% | 89.4% |
| | Prev. | 6.20 | 15.50 | | | | 0.1% | 11.7% | 88.1% |
| Offside Rut | Curr. | 5.10 | 14.80 | 0.99 | 0.17 | -0.54 to 0.54 | 0.1% | 7.9% | 92.0% |
| | Prev. | 4.93 | 14.80 | | | | 0.1% | 8.1% | 91.8% |
| Nearside 3m LPV | Curr. | 3.07 | 28.10 | 0.99 | -0.43 | -1.3 to 1.30 | 4.6% | 14.1% | 81.3% |
| | Prev. | 3.50 | 31.80 | | | | 5.8% | 16.6% | 77.6% |
| Nearside 10m LPV | Curr. | 27.28 | 342.87 | 0.99 | -3.81 | -2.5 to 2.50 | 13.0% | 24.8% | 62.2% |
| | Prev. | 31.09 | 440.63 | | | | 16.7% | 27.7% | 55.5% |
| Nearside SMTD | Curr. | 0.76 | 0.23 | 1.00 | -0.02 | -0.07 to 0.07 | 4.2% | 29.6% | 66.3% |
| | Prev. | 0.79 | 0.24 | | | | 3.2% | 27.7% | 69.0% |
| Cracking | Curr. | 0.31 | 3.00 | 1.00 | 0.01 | -0.18 to 0.18 | 3.4% | 32.6% | 64.0% |
| | Prev. | 0.29 | 2.90 | | | | 2.9% | 32.0% | 65.1% |
| RCI | Curr. | 39.88 | 150.00 | 1.00 | 1.10 | -12.88 to 12.88 | 7.6% | 36.0% | 56.4% |
| | Prev. | 38.78 | 145.61 | | | | 7.0% | 35.6% | 57.4% |

| Audit Indicator[4] | | | Expected Range of Indicator for Curr. year |
|---|---|---|---|
| | Curr. | 7.6% | |
| | Prev. | 7.0% | 5.3% to 8.7% |

[1] 1st Percentile for NS SMTD [2] Using RCI thresholds for urban B roads [4] Percentage of RCI values ≥ 100 (**Red** Category)
[3] Red = Max contribution to the RCI, Amber = Contributing to the RCI but not maximum, Green = Not contributing at all

**2 of 3 audit criteria met**

**Notes & Suggested Actions:**
Bias in 10m LPV outside expected range.

For further information please contact your current SCANNER survey provider.
If you then require additional information please email scanner@trl.co.uk.

Only an issue with 10m LPV bias for Herefordshire




## Task 2: Consistency of eLPV – QA Herefordshire Audit Report, using eLPV

LINHAY CONSULTANCY hyperion infrastructure consultancy

| Param. | Data | Mean | 99th percentile[1] | Distn. Corr. | Bias from Prev. | Expect Bias Range | Red[3] | Amber[3] | Green[3] |
|---|---|---|---|---|---|---|---|---|---|
| Nearside Rut | Curr. | 5.61 | 16.10 | 0.96 | -0.60 | -0.71 to 0.71 | 0.2% | 10.4% | 89.4% |
| | Prev. | 6.20 | 15.50 | | | | 0.1% | 11.7% | 88.1% |
| Offside Rut | Curr. | 5.10 | 14.80 | 0.99 | 0.17 | -0.54 to 0.54 | 0.1% | 7.9% | 92.0% |
| | Prev. | 4.93 | 14.80 | | | | 0.1% | 8.1% | 91.8% |
| Nearside 3m LPV | Curr. | 1.50 | 15.44 | 0.99 | -0.20 | -1.24 to 1.24 | 4.1% | 12.4% | 83.5% |
| | Prev. | 1.70 | 17.54 | | | | 5.2% | 14.3% | 80.6% |
| Nearside 10m LPV | Curr. | 6.81 | 57.87 | 0.99 | -1.06 | -2.2 to 2.20 | 4.5% | 14.9% | 80.7% |
| | Prev. | 7.87 | 66.55 | | | | 5.8% | 17.8% | 76.4% |
| Nearside SMTD | Curr. | 0.76 | 0.23 | 1.00 | -0.02 | -0.07 to 0.07 | 4.2% | 29.6% | 66.3% |
| | Prev. | 0.79 | 0.24 | | | | 3.2% | 27.7% | 69.0% |
| Cracking | Curr. | 0.31 | 3.00 | 1.00 | 0.01 | -0.18 to 0.18 | 3.4% | 32.6% | 64.0% |
| | Prev. | 0.29 | 2.90 | | | | 2.9% | 32.0% | 65.1% |
| RCI | Curr. | 34.47 | 142.80 | 1.00 | 1.90 | -12.88 to 12.88 | 5.7% | 32.1% | 62.2% |
| | Prev. | 32.57 | 140.00 | | | | 5.1% | 30.5% | 64.5% |

| Audit Indicator[4] | | | Expected Range of Indicator for Curr. year |
|---|---|---|---|
| | Curr. | 5.7% | |
| | Prev. | 5.1% | 3.5% to 6.6% |

[1] 1st Percentile for NS SMTD [2] Using RCI thresholds for urban B roads [4] Percentage of RCI values ≥ 100 (**Red** Category)
[3] Red = Max contribution to the RCI, Amber = Contributing to the RCI but not maximum, Green = Not contributing at all

**All 3 audit criteria met**

**Notes & Suggested Actions:**

For further information please contact your current SCANNER survey provider.
If you then require additional information please email scanner@trl.co.uk.

Using eLPV removes this issue




## Task 2: Investigating the step change in the RCI caused by eLPV

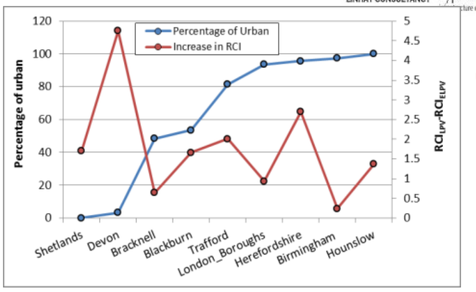
LINHAY CONSULTANCY hyperion infrastructure consultancy

- Calculated change in Audit Indicator (percentage of red lengths for one year's data) for several authorities:
  - Shetlands, Herefordshire? and Devon (very rural authorities);
  - Bracknell and Blackburn (semi-rural authorities);
  - Trafford (metropolitan authority)
  - London boroughs
  - Birmingham and Hounslow (urban authorities).
- Gives an indication of the effect on the NIs




## Task 2: Investigating the step change in the RCI caused by eLPV

LINHAY CONSULTANCY hyperion infrastructure consultancy




Change in audit indicator loosely proportional to the amount of rural lengths. Approximately 1.5% decrease for most authorities.




## Task 2: Including offside data in RCI

LINHAY CONSULTANCY hyperion infrastructure consultancy

- Contribution from rutting is the maximum from the nearside and offside
- Contribution from LPV is just from nearside but also have offside data
- What is the effect of using max (offside and nearside LPV or eLPV) in RCI?
- Working on this.



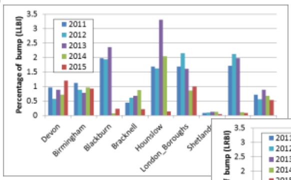

## Task 2: Investigation of Bump Measure

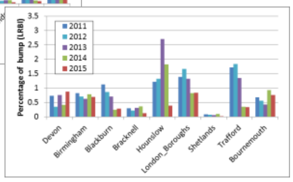
LINHAY CONSULTANCY hyperion infrastructure consultancy

- Bump not used by engineers, despite there being a desire for a measure of potholes in SCANNER;
- Bump thought to be inconsistent (on a length by length basis) and affected by driving line;
- Investigate whether it could be used to provide a network level indicator of the presence of potholes and bump-causing features.

## Task 2: Investigation of Bump Measure – network level

LINHAY CONSULTANCY hyperion infrastructure consultancy




For some authorities, the percentage of network affected by bump is fairly constant. For others it's not…

Splitting by road class doesn't change this result.

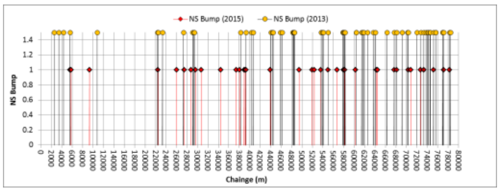


Page • 71

## Task 2: Investigation of Bump Measure – analysis on individual lengths

LINHAY CONSULTANCY hyperion infrastructure consultancy

High level investigation of why bump measure is inconsistent, using 2013 and 2015 data from Bournemouth.




Length by length comparison would suggest that Bump Measure is not good



Page • 72

## Task 2: Investigation of Bump Measure – analysis of individual lengths

LINHAY CONSULTANCY hyperion infrastructure consultancy

However, there's always a bump causing feature where the Bump Measure has triggered

Yellow pins are 2013 data, Blue are 2015. Red pins are where the bump measure has triggered



Page • 73

## Task 2: Investigation of Bump Measure – analysis of individual lengths

LINHAY CONSULTANCY hyperion infrastructure consultancy

Bump Measure triggered in both years for one length, only one year for other.



Page • 74

## Task 2: Investigation of Bump Measure – analysis of individual lengths

LINHAY CONSULTANCY hyperion infrastructure consultancy

Length where Bump Measure triggered in only one year. Feature is very localised



Page • 75

## Task 2: Investigation of Bump Measure – analysis of individual lengths

LINHAY CONSULTANCY hyperion infrastructure consultancy

Transverse feature on length where Bump Measure triggered in both years



Page • 76

## Task 2 – Summary

LINHAY CONSULTANCY hyperion infrastructure consultancy

1. Likely improvement if eLPV replaces LPV
   a. Less affected by geometry
   b. As or more consistent
   c. Reduction of ~1.5% in Audit Indicator when using eLPV to calculate RCI
   d. Need to investigate step change caused by including offside ride quality in the RCI calculation (likely to cause increase).
2. Bump measure
   a. Does identify real features
   b. Not consistent year on year
   c. Not consistent between devices
   d. Need to confirm whether this is because of driving line
   e. Investigate whether it can be improved.



Page • 77

## Task 2 – Next steps

LINHAY CONSULTANCY hyperion infrastructure consultancy

1. Investigate step change to audit indicator when including offside ride quality in RCI
2. Confirm whether driving line causes inconsistency in Bump and investigate what might be done to improve the measure, or provide an alternative measure within SCANNER
3. Consider multiple line texture and how measure of fretting might be developed
4. Investigate the capability of SCANNER to provide a better bump/pothole measure.



Page • 78

# Development of SCANNER and UKPMS: Task 1 - Consistency of SCANNER data and Task 2 - SCANNER Condition Parameters

## Other titles from this subject area

| | |
|---|---|
| **PPR 817** | "Development of SCANNER and UKPMS: Task 3 - Appropriateness of the SCANNER RCI". CC Spong (Hyperion Infrastructure Consultancy) and R A Cartwright (Linhay Consultancy). 2017 |
| **PPR014** | "Initial study and development of transverse profile analysis – TTS on local roads". K Nesnas, S McRobbie & A Wright. 2004 |
| **PPR131** | "Shape (surface form) of local roads". E Benbow, K Nesnas & A Wright. 2006 |

**TRL**
Crowthorne House, Nine Mile Ride,
Wokingham, Berkshire, RG40 3GA,
United Kingdom
T: +44 (0) 1344 773131
F: +44 (0) 1344 770356
E: enquiries@trl.co.uk
W: www.trl.co.uk

ISSN

ISBN

**PPR816**