# Improvement of visual road condition data
June 2013

MSP Tapper, Beca, Tauranga
TFP Henning, University of Auckland
KD Dunn, Beca, Tauranga

**NZ Transport Agency research report 528**

ISBN 978-0-478-40761-7 (electronic)
ISSN 1173-3764 (electronic)

NZ Transport Agency
Private Bag 6995, Wellington 6141, New Zealand
Telephone 64 4 894 5400; facsimile 64 4 894 6100
research@nzta.govt.nz
www.nzta.govt.nz

Beca Infrastructure was contracted by NZTA in 2010 to carry out this research

Tapper, MSP, TFP Henning and KD Dunn (2013) Improvement of visual road condition data. *NZ Transport Agency research report 528.* 72pp.

**Keywords:** asset management, data collection, pavement condition rating, surface distress, visual condition survey

# An important note for the reader

The NZ Transport Agency is a Crown entity established under the Land Transport Management Act 2003. The objective of the Agency is to undertake its functions in a way that contributes to an affordable, integrated, safe, responsive and sustainable land transport system. Each year, the NZ Transport Agency funds innovative and relevant research that contributes to this objective.

The views expressed in research reports are the outcomes of the independent research, and should not be regarded as being the opinion or responsibility of the NZ Transport Agency. The material contained in the reports should not be construed in any way as policy adopted by the NZ Transport Agency or indeed any agency of the NZ Government. The reports may, however, be used by NZ Government agencies as a reference in the development of policy.

While research reports are believed to be correct at the time of their preparation, the NZ Transport Agency and agents involved in their preparation and publication do not accept any liability for use of the research. People using the research, whether directly or indirectly, should apply and rely on their own skill and judgement. They should not rely on the contents of the research reports in isolation from other sources of advice and information. If necessary, they should seek appropriate legal or other expert advice.

# Acknowledgements

NZTA is recognised for its funding of the research project. Also the full research team and steering group are acknowledged for their respective contribution towards this research. They are:

Research Team: A. Nataadmadja, S. Christy, D. Nicholson and T. Allan

Steering Group: M Hendry (Research Manager), J Brass, G Morrow, C Thew, J Chisnall. M McMinn

# Abbreviations and acronyms

| | |
|---|---|
| CI | Condition Index |
| dTIMS | Deightons Total Infrastructure Management System |
| FWP | forward works programme |
| HDM | highway design and maintenance standards model |
| HSDC | high-speed data capture |
| IRI | International Roughness Index |
| KPI | key performance indicator |
| L&T | longitudinal and transverse |
| LTPP | long-term pavement performance |
| LWP | left-wheel path |
| NZ | New Zealand |
| NZIHT | New Zealand Institute of Highway Technology |
| NZTA | New Zealand Transport Agency |
| PFM 6 | *RAMM road condition rating and roughness manual* |
| PII | Pavement Integrity Index |
| QA | quality assurance |
| RAMM | road asset and maintenance management |
| RCA | road controlling authority |
| SCI | Surface Condition Index |
| SCRIM | sideway-force coefficient routine investigation machine |
| SH | state highway |
| SII | Surface Integrity Index |
| SWC | surface water channel |
| TLA | territorial local authority |
| TSA | treatment selection algorithm |
| TSF | thin surface flexible |

# Contents

# Executive summary

The aim of this research project, undertaken in 2011, was to identify recommendations to the current *RAMM Road rating and roughness manual* to improve data accuracy and consistency.

Originally, the purpose of the visual condition survey data was to run the road asset and maintenance management (RAMM) treatment selection algorithm analysis. The results would provide a list of candidate sections to be validated in the field forming a maintenance works programme of resurfacings and rehabilitations. However the condition data is now also used to feed into key performance measures (KPI) and levels of service measures. The NZTA, furthermore, wishes to extend the use of these KPIs to compare the performance of different road controlling authority (RCA) networks.

Currently the visual condition rating data is used to describe or analyse pavement and surfacing condition in the following:

- Surface Condition Index (SCI)
- Pavement Integrity Index (PII)
- RAMM treatment selection algorithm (TSA) analysis
- NZdTIMS modelling
- condition trends.

Thus the role of visual condition rating, particularly with the advent of the RAMM hosting server and pavement deterioration modelling, is significantly different from when it was first developed. The process therefore needs to be improved to better reflect the purposes for which it is currently used.

## Data collection review

For alligator cracking the variability of rater values during the annual rating course was found to be too high. The quality of data recorded for this parameter is important for having confidence in the results of the performance measures, the TSA and modelling. A visual rating survey is currently the best method for identifying this fault type; however, measures need to be established to improve accuracy and confidence.

The current acceptable limits of variation need to be tightened to avoid under or over-reporting of this fault type. Current limits allow variations in reporting that result in differences in treatments predicted in the RAMM TSA results for example. The *RAMM road condition rating and roughness manual* should include better examples. More emphasis at the annual rating course should be focused on cracking as well as improved quality assurance (QA) procedures.

Rutting is increasingly used as a measure of pavement performance. Currently it is rated by reporting the length of wheelpath rut depth greater than 30mm. However high-speed data surveys report a measure of average rut depth in each wheelpath. This form is much more useful for modelling purposes and the continuous data stream allows greater statistical analysis of the distribution of rut depth.

It is recommended that the rutting data be collected visually by assessing the length of the wheelpath to the nearest 5mm, similar to the alternative method currently in the rating manual. This would create a methodology consistent with high-speed data capture and create less of an issue for comparing network condition and indices where different data collection methods are used.

The manual requires updating with improved guidance on fault definition and should include photographs of fault types, particularly of different types of cracking.

## Rater training

The current approved road condition rating course is run annually by the New Zealand Institute of Highway Technology. The accreditation course is a two-day course for new raters. Accredited raters attend the second day only as a refresher workshop every two years. There is an assessment but the limits of variation are such that it is very difficult to fail. It is recommended that the assessment criteria be tightened so that raters leave the course with a consistent approach to rating. The opportunity for feedback to raters may need to be improved, perhaps through limiting numbers to allow better individual assessment and feedback.

It should be noted, however, that the consistency of survey results should be achieved through the application of appropriate and effective QA systems during the surveys in the field.

## Quality assurance procedures

The *RAMM road condition rating and roughness manual* currently requires a 5% validation area to ensure accuracy of the data collected. It is recommended the manual is improved to include QA practice guidelines on the implementation of a common rating sample which is surveyed by all raters, including the QA person, to identify any consistent under or over-reporting. An assessment will need to be made on the level of variance that requires intervention. The limits of variation would not need to apply. This could be a measure based on standard deviation from the mean for each rater

A second major recommendation is the implementation of an independent validation service to achieve a better consistency across networks, for example, an NZTA funded audit team or an alternative supplier to do a sample on certain networks or road hierarchies of key importance.

## Stratification and sampling

A maximum sample length of 200m is recommended. A minimum 20m inspection length would provide a 10% sample while on higher volume roads, a minimum of 40m or 20% can be used. RCAs can still choose to go to 100% sampling to provide greater confidence and accuracy. However the impact of underreporting faults will be minimised. This stratification could be undertaken by either traffic volume (eg greater or less than 500 vpd) or by hierarchy (say local roads at 10%, arterials, collectors, etc at 20%).

The use of 200m sections would not require any split between urban and rural road sections simplifying the autorate process of generating forms.

It is recommended that RCAs undertake condition rating surveys at a consistent time of year on their network. This will aid with consistency when analysing trend data, etc.

The use of high-speed data capture surveys on higher level roads is recommended.

## Procurement

We recommend the following for stand-alone condition assessment contracts:

- multi-year, preferably 3+1+1 term contracts, or 4+2 if roads are surveyed on alternate years
- weighted attribute rather than lowest price conforming
- specified QA requirements, preferably based on best practice guidelines as part of an updated rating manual
- a single combined contract for smaller networks, similar to the system operated successfully by Hurunui, Waimakariri and Kaikoura districts for a number of years.

Documentation is consistent in our experience although QA practice can vary. However, the documentation is generally standard, particularly concerning rating requirements, deliverables, and limits of variation and calibration procedures. An improved guide to QA procedures in the rating manual will assist with this.

# Abstract

The objective of this research, which was carried out between 2010 and 2012, was to investigate the effectiveness of the current road condition rating system with a view to improving the accuracy and confidence in the data collected. This in turn will build confidence in key network performance indicators.

The use of visual road condition rating data in New Zealand has evolved from its original purpose of identifying carriageway sections on a network level for treatment and from being employed in the development of a forward works programme. Visual rating data is now used as an input into a series of performance measures and other pavement/surfacing performance modelling. This research project looked at the how the visual rating process is currently undertaken and whether this is appropriate for its current and future uses. With the move towards using the data to compare road controlling authority networks, confidence and consistency in the data is paramount.

The research recommends improvements to data collection methodology, rater training, quality auditing, survey stratification and sampling methodology and procurement.

# 1 Introduction

The use of visual road condition rating data in New Zealand has evolved from its original purpose of identifying carriageway sections on a network level for treatment, and subsequently contributing to the development of a forward works programme (FWP). Visual rating data is now used as an input into a series of performance measures and other pavement/surfacing performance modelling.

This research project looked at how the visual rating process is currently undertaken and whether the process is appropriate for its current and future uses. With the move towards using the data to compare road controlling authority (RCA) networks, confidence and consistency in the data is paramount. The use of advancing technology through high-speed data capture (HSDC) surveys was also considered.

Visual road rating is currently undertaken to identify pavement and surfacing defect types indicating condition. The *RAMM road condition rating and roughness manual* (PFM 6) (Transfund NZ 1997) provides guidance on undertaking the visual rating. It also sets out the acceptable limits of variation (tolerance) allowed on the defect value recorded when checked through a quality assurance (QA) audit process.

Different RCAs undertake the rating with varying sample sizes. For rural and state highway networks the rating pattern is generally 10% of the network, ie rating 50m every 500m. For urban networks, up to 100% of the network is rated with the sections tending to be shorter as a result of the network layout.

Annual road rater training certification courses are currently run by the New Zealand Institute of Highway Technology (NZIHT). These aim to provide consistency in identifying and measuring defects by raters throughout New Zealand. The course is run over two days for new raters. The first day is split between classroom and site learning with the trainer. The second day is an assessment where attendees rate a number of sites and the results are compared with those produced by the trainers. All raters must attend a refresher of this training every two years which consists of the second day.

There is currently no industry guidance into methodologies to provide QA in the data. As a result this varies from contract to contract, and is either clearly identified by the client, left to the contractor or not mentioned at all.

The aim of this research, undertaken in 2011, was to identify recommendations for improving the data accuracy and consistency of the current PFM 6.

Work was undertaken as part of this research to establish:

- the impact of the rating data on current and future key performance indicators (KPIs), levels of service and forward work programming
- the appropriateness of the guidance given in the PFM 6 to achieve data accuracy and consistency
- the effects of the sampling regime on data accuracy and consistency including sample size, time of year and speed of survey
- the effectiveness of the annual rating course
- methodologies to provide QA in the data.

A number of recommendations were identified to improve the accuracy and consistency of the road condition rating data including an assessment of the impact on survey costs.

# 2 Literature review

## 2.1 Purpose and scope of the literature review

The main purpose of the literature review was to investigate practices around known issues with the current visual rating scheme used in New Zealand. Therefore it was by no means a complete academic literature review covering all aspects of the topic area. It was outside the scope of the research project to look at alternative rating systems. But we sought to identify any lessons learnt or possible improvements to the current system.

The scope of the literature review included:

- survey methods
- defect types surveyed
- sampling methods
- frequencies of surveys
- timing of surveys
- QA processes.

The following sections discuss these topics in more detail.

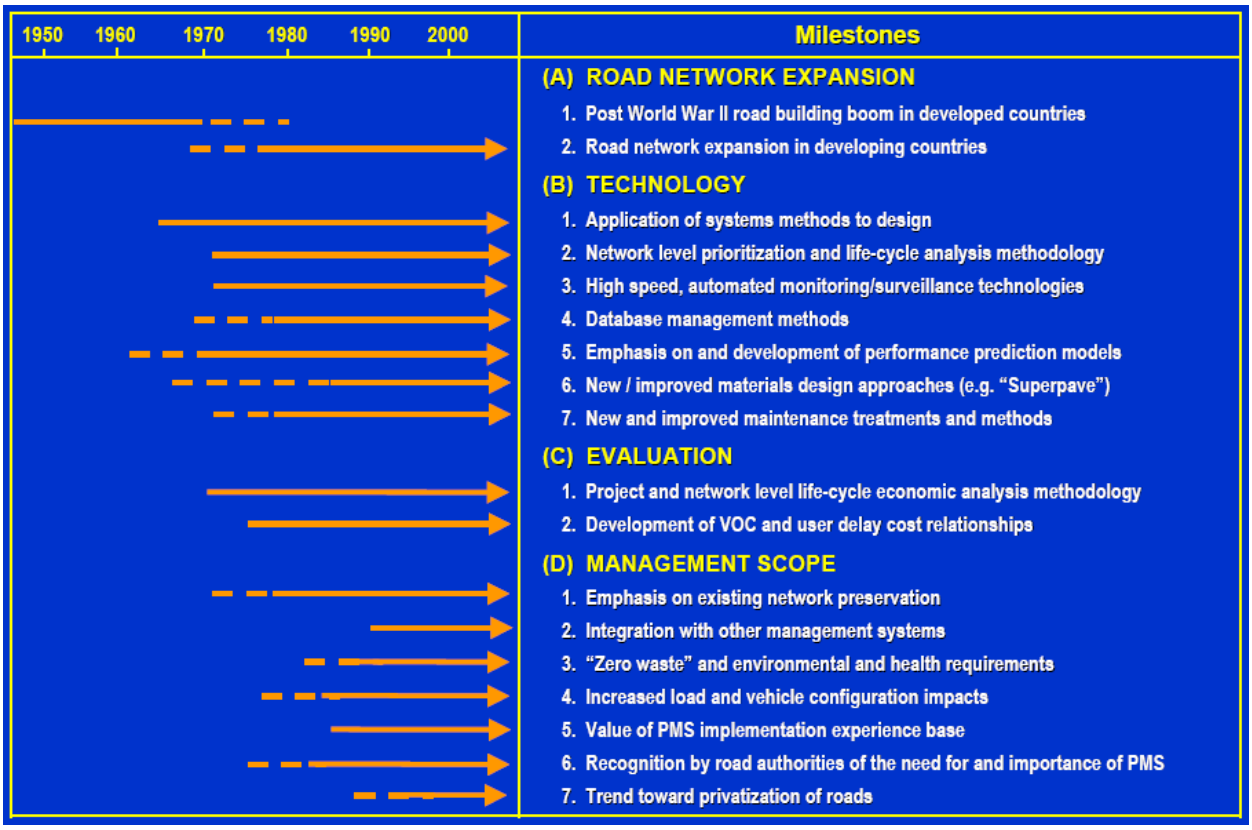
## 2.2 The purpose and methods of road data collection

### 2.2.1 The development history

Since the early 1970s there has been a significant transition from subjective assessment of road conditions to more automated processes. As illustrated in figure 2.1, not only did we notice a significant increase in technological sophistication in the development of data collection techniques, but there has also been a significant development in the actual use of the data. For example, originally surveys were undertaken mainly to schedule maintenance work, whereas they are now used for advanced performance monitoring and forecasted work programmes.

Likewise, the road asset and management maintenance (RAMM) survey methodology was originally adopted for qualifying road condition and as an input into a short-term decision algorithm (treatment selection algorithm). In 1999, with the implementation of the Deightons Total Infrastructure Management System (dTIMS) RAMM data was used as an input into the World Bank HDM-III prediction models. For this, a conversion process was adopted to change the RAMM assessment scale (length of wheel path affected) to a percentage of carriageway affected (HTC 2000). The NZTA has been using condition trend information from local authorities since 2003 to test the distribution of maintenance funds across the country. In a recent rating review, Pradhan (2009) summarised the various uses of the RAMM condition data by a number of councils. The result from his survey is depicted in figure 2.2. It shows that most authorities use the RAMM data for an array of asset management applications. It is therefore fair to conclude that the RAMM survey data has been used well outside the original scope for which it was developed.

Figure 2.1 Development time-line of pavement management systems (Haas 2001)



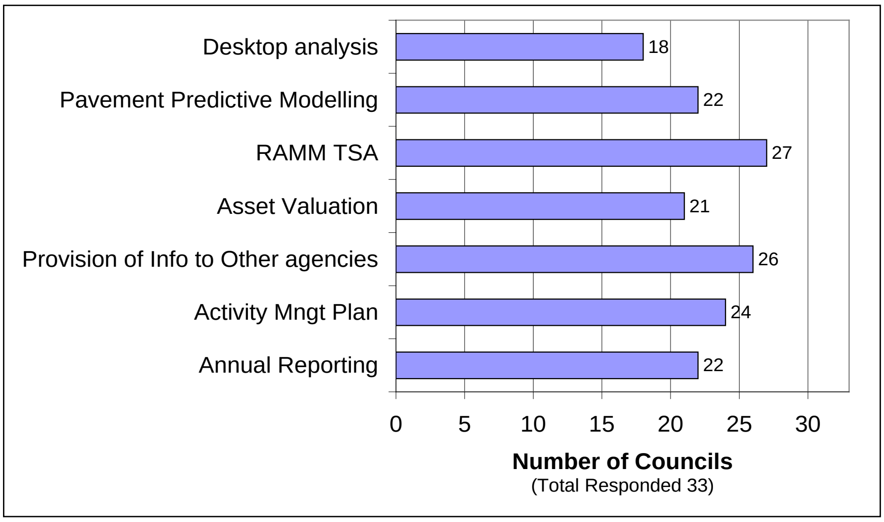
(A) ROAD NETWORK EXPANSION
1. Post World War II road building boom in developed countries
2. Road network expansion in developing countries

(B) TECHNOLOGY
1. Application of systems methods to design
2. Network level prioritization and life-cycle analysis methodology
3. High speed, automated monitoring/surveillance technologies
4. Database management methods
5. Emphasis on and development of performance prediction models
6. New / improved materials design approaches (e.g. "Superpave")
7. New and improved maintenance treatments and methods

(C) EVALUATION
1. Project and network level life-cycle economic analysis methodology
2. Development of VOC and user delay cost relationships

(D) MANAGEMENT SCOPE
1. Emphasis on existing network preservation
2. Integration with other management systems
3. "Zero waste" and environmental and health requirements
4. Increased load and vehicle configuration impacts
5. Value of PMS implementation experience base
6. Recognition by road authorities of the need for and importance of PMS
7. Trend toward privatization of roads

Figure 2.2 The use of rating data in New Zealand (Pradhan 2009)



With the increased focus on using pavement condition data on a network reporting and analysis level, there has been an emphatic shift away from manual surveys to more automated surveys. For example, the state highways are now using the SCRIM+ machine for capturing a number of condition items. All reporting and trend analyses of state highways rely mostly on automated data with the only exception being the Surface Integrity Index which uses visually rated data for faults such as cracking (NZTA 2009).

The trend of switching from manual to automated data has not always been maintained and some cities, such as Denver, are reverting to manual assessment of road ride quality (Piane 2010). However, in some areas such as motorways, scanning lasers and video images are now being used for assessing condition items such as cracking, where traditionally these defects were assessed using visual methods.

It can be safely assumed that in New Zealand visual condition rating will continue for the foreseen future, mostly because of the extensive coverage of thin, flexible, chip seal pavements. Visual condition items such as cracking and ravelling need to be identified early, something which is currently only possible by visual assessment.

### 2.2.2 Visual rating methods

Visual rating methods can be classified into three categories including:

1. Windshield rating – with this method, the rater drives along at a slow speed (typically 20km/h). Faults are recorded on an inspection form or by electronic interfaces. The rating is used to assess 100% of the network but given the speed of the surveys, it is accepted that it is not extremely accurate. A typical 5-point scale is used to assess the degree and extent of defects for an entire road section (CSRA 1992).

2. Manual rating on foot, such as the NZ RAMM method, where detailed recording of defects is undertaken for a small sample of the treatment length (typically 10%). For this assessment the degree of the defect is not rated but a detailed extent estimation (length of wheelpath affected) is recorded.

3. Analysis of electronic images – this method relies on either still images captured at high speed or video images to assess 100% of the network. This is similar to the windshield type assessments with the only difference being that the rating takes place in office conditions (Fwa et al 2003).

Some literature compares the merits of these methods. Ultimately, the aim is to achieve an acceptable level of accuracy for the intended purpose of the condition data, taking into consideration resource constraints. For example, if high accuracy is required in the data collection, a manual rating process is necessary, but 100% network coverage then becomes more expensive. It is believed though that for the New Zealand conditions, environment and type of pavements, a manual visual rating system is still appropriate.

## 2.3 Known issues defect types surveyed

Given the nature of flexible, thin-surfaced pavements, more defects are assessed on these pavements compared with deep-lift structural asphalt pavements. Currently the RAMM survey method allows for 11 different carriageway distress types. These are:

- alligator cracking
- longitudinal and transverse cracking
- scabbing
- edgebreak
- shoving
- joint cracking
- potholes
- edgebreak patches
- rutting
- flushing
- pothole patches

There are a number of known issues that Pradhan (2009) has identified and these are listed in table 2.1.

**Table 2.1 Known issues with RAMM rating defects (Pradhan 2009)**

| Issues raised | Potential solutions suggested |
|---|---|
| Cracking is interpreted differently by different raters. | More specific definition and illustrations of various cracking need to be included in new manual.<br>Differentiate between the superficial and structural cracking types. |
| Wide variation in the rating of scabbing by different raters is affecting historical data and national key performance indicator trend analysis. | More specific definition and illustrations of various scabbing to be included in the new manual. |
| Defects related to trench patches, service covers not fully assessed. | Consider a future NZTA research study into data collection of trench patches and service cover defects |
| Pothole patches are too small to cover dig out works. | Include three different size of pothole/digout patches in the new manual. |
| Rutting data is difficult to collect and not reliable. | Consider using a profilometer for rutting data collection at defined intervals (eg every 3 years). |
| Too many parameters are included in the existing visual condition rating procedure, hence, time consuming and costly. | Need to carry out a study to find which measurements are not used and consider making them redundant.<br>Change methodology of rating, eg give global rating of section and identify two major defects. |

Some of the issues identified are discussed further in subsequent sections.

### 2.3.1 Scabbing and flushing

During the development of the New Zealand long-term pavement performance (LTPP) survey specification, detailed requirements were developed for identifying all signs of visual distress (Transit NZ 2000). For most distress types, a detailed recording of size and location was required. However, both scabbing and flushing were quantified by degree and extent measures. For these distress types there is an element of:

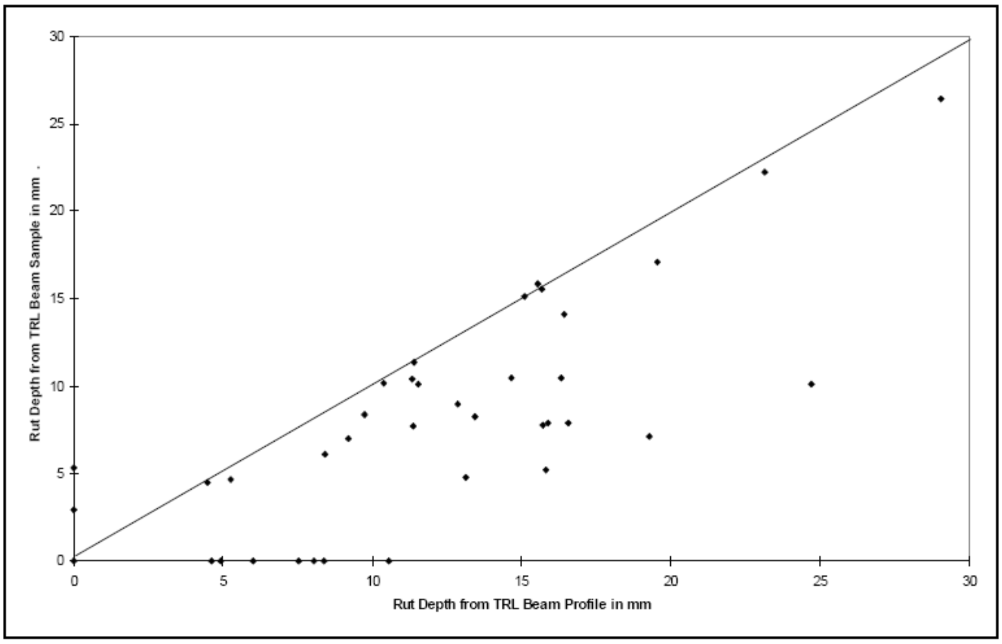
- how serious the defect is
- how widely a rated section is affected by it.

It was recommended that this research explored ways of assigning measures of degree and extent to these defects.

### 2.3.2 Rutting

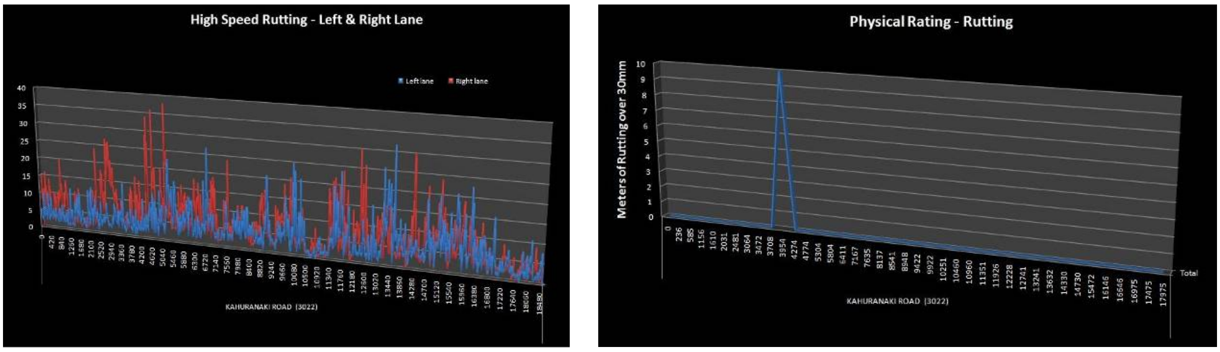
There are a number of studies that suggest poor information is provided through the RAMM ratings of rutting. For example, Bennett (2001) demonstrated the effectiveness of using sampled rutting compared with a continuous measurement (see figures 2.3 and 2.4).

**Figure 2.3 Comparing sampled rut depth with continuously measured rut depth (Bennett 2001)**



Note that this figure compares high-speed rutting data using continuous measures versus using the data by means of sampling. HDC (2013) also considered only rutting from HSD measures but demonstrated the value of only considering rutting above 30mm.

**Figure 2.4 Reporting continuous versus rutting above 30mm (Thew 2009)**



Continuous reporting of rutting

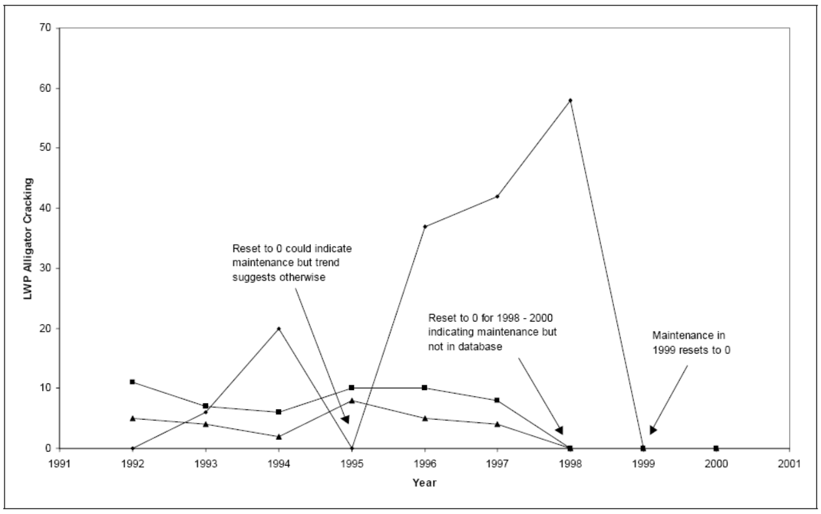
Exception reporting of rutting

This figure shows the skewed conclusions that could be drawn from considering rutting above a certain level as an indication of rutting progression. Based on this evidence it is recommended that manual rut surveys be deleted from the New Zealand rating method. Alternatively, the rating method should be adjusted to provide a ‘condition and extent’ measure or utilise a methodology that mirrors that of the HSD survey. An example would be reporting lengths within 5mm deep bandwidths.

## 2.3.3 Cracking

Various research (Henning et al 2006; Pradhan 2009; Perera 2010) has demonstrated the poor trends that can be derived from information on the extent of cracking. As illustrated in figure 2.5, there is a number of reasons and issues that lead to poor historical crack information. These include referencing issues, poor recording of maintenance history, and confusion by raters about the type of cracking they are assessing.

**Figure 2.5 Poor trends from crack information (Bennett 2001)**



However, alligator cracking is one of the most important drivers in maintenance decisions, thus it should be captured by some means. Also, it is not always the extent of cracking that has the significant bearing on a maintenance decision. Henning (2009) demonstrated that relatively strong trends could be observed from information about the outset of cracking. Therefore, apart from investigating ways of improving the quality of crack information, consideration could also be given to creating database fields that record the first occurrence of cracking.

## 2.4 Sampling methods

As indicated earlier, it is believed that the most appropriate method for New Zealand conditions is to have more accurate surveys based on a sampling approach. Some authorities have found that a 10% sample is not sufficiently accurate and there is widespread belief that the sample size should be increased. However, no literature could be found that suggested an appropriate level of sampling for road condition measurements.

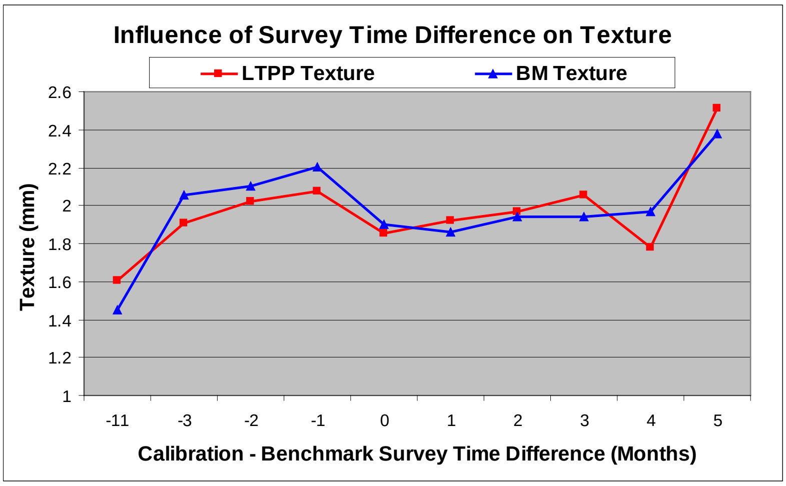
It was thus recommended that this research establish an appropriate sampling size for New Zealand rating methods. The sampling size should be established from appropriate confidence levels needed for trend analysis and deterioration modelling.

In addition to the sampling size issue, the research also had to consider ways of achieving greater consistency in the rating length used for surveys. It is acknowledged that treatment lengths change significantly over time. However, a process needed to be developed that would keep rating sections consistent.

## 2.5 Frequencies and timing of surveys

It is a well-established fact that timing has a significant impact on condition surveys. In their study Deng and Henning (2013) found that timing of surveys was one of the factors that caused the most variation in survey outcomes. For example figure 2.6 illustrates the influence time has on the repeatability of condition surveys. The figure shows the influence of the timing of surveys on texture measurement. The time difference is the number of months between the LTPP calibration survey and the benchmark survey. This figure shows a significant difference in relation to the timing of surveys. Although this trend is most evident for texture, it also suggests an automatic influence on cracking given the self-healing phenomena of bitumen during hot climatic conditions.

**Figure 2.6 Difference in condition surveys as a functioning of survey timing (Deng and Henning 2013)**



Therefore, regardless of the frequency of surveys, we recommend stricter guidelines on the timing of surveys, which should take place within a specified period of the survey year. This is common practice in other countries such as South Africa.

The current requirement for authorities is to carry out specified RAMM surveys covering a network over a two to three-year period. Although the NZTA's (2011) Planning and Investment Knowledge Base agrees that surveys do not have to be undertaken every year (depending on traffic volumes), surveying only parts of the network annually leads to significant inconsistencies in condition trends. We therefore recommend that specifications still require a survey to be undertaken annually or bi-annually, but that complete networks are surveyed during the rating processes.

## 2.6 Quality assurance processes

Current rating methods lack QA processes. These should consist of two components:

1 Ensuring raters are sufficiently trained and competent for the tasks at hand

2 Testing survey quality and repeatability to ensure the quality of the surveys is sustained.

The above measures may lead to an increase in the unit cost of surveys, but the value returned from these surveys should far outweigh any cost increases.

## 2.7 Summary of literature review findings

A summary of the literature findings with recommended further work is presented in table 2.2.

**Table 2.2 Summary of literature findings and further work**

| Item | Findings | Further work |
|---|---|---|
| Purpose of surveys | The purpose of the RAMM rating has changed significantly over time. | Re-define the current and likely future use of rating information, with particular emphasis on accuracy requirements. |
| Rating methods | Different rating methods exist but it is believed that the current approach is appropriate for New Zealand application. | Policy decision – confirm accuracy requirements through this research. |
| Cracking recording | Clearer instructions for crack types are required. | Guidelines to include more background to the causes of crack types.<br>Data management to include recording crack initiation time. |
| Scabbing and flushing | Current information is not consistent and sufficient for assessing seriousness of defects. | Devise a mechanism to include an indication of degree and extent. |
| Rutting | Very poor results are obtained from current rutting ratings. | Discontinue as rated item or investigate alternative rating methodologies. |
| Sampling methods | A sampling method is still applicable but some measures of consistency are required. | Test sampling size as a function of confidence levels required.<br>Review rating section generation to ensure more consistency for trend reporting. |
| Frequency and timing of surveys | Timing of surveys needs to be more consistent.<br>Annual surveys would not be required but when surveys take place they should cover the total network. | Investigate as part of this research. |
| Quality assurance process | All processes related to QA need reviewing. | Review as part of this research. |

# 3 Impact of data collection parameters

## 3.1 What is visual condition data used for?

The original purpose of the visual condition survey data was to run the RAMM treatment selection algorithm analysis. This provides a list of candidate sections to be validated in the field and forms a maintenance works programme of resurfacings and rehabilitations. However the condition data is now also used to feed into KPI and levels of service measures. The NZTA, furthermore, wishes to extend the use of these KPIs to compare the performance of different RCA networks.

This is a key shift for the use of the condition data, from primarily an internal network use, to one of comparing data between networks. Therefore, the consistency and confidence of relative data is more important. Trend analyses using the data are much more prevalent given the wider availability of the data through the hosted RAMM server and the improved use of technology. Also factors such as the impact of the time of year the data is collected and differing sampling regimes become more crucial.

There are six key uses and/or issues for data now:

- Running the treatment selection algorithm (TSA) analysis in RAMM. The outputs produce a list of candidate road sections to be validated in the field, forming a maintenance works programme of resurfacings and rehabilitations.
- Internal RCA network condition trend analyses and desktop analysis.
- RCA network reporting and input into internal documents such as asset and activity management plans, asset valuation and annual reporting.
- Comparison between peer groups of networks and national data. The condition data is used to feed into KPIs and level of service measures.
- As a key component in the use of pavement deteriorating models such as the dTIMS.
- The NZTA seeks to extend the use of data within KPIs to compare the performance of different RCA networks and therefore assess maintenance needs.

The last three bullet points reflect the new comparative usage of the data. This has created a new emphasis on providing comparable and consistent data between networks. It has also created a new requirement for understanding the level of confidence in the data accuracy, particularly in using the data to assess and prioritise maintenance needs on a national basis.

## 3.2 Which parameters are used to describe pavement and surface condition?

Currently the visual condition rating data is used to describe or analyse pavement and surfacing condition in the following:

- Surface Condition Index (SCI)

- Pavement Integrity Index (PII)
- RAMM treatment selection algorithm (TSA) analysis
- NZdTIMS modelling
- condition trends.

The first two, SCI and PII, are indices used by the NZTA to describe pavement and surfacing condition for RCA road networks. These are used in the standard NZTA annual reporting. TSA is used by RCAs to assess candidate sections for treatment. The NZdTIMS pavement deterioration modelling software also uses visual road condition data to predict pavement performance.

The outcomes from the five aspects listed above are driven by the carriageway faults recorded during the visual road rating. Surface water channel faults are only used as an input into TSA as part of the economic analysis and likely future resurfacing cycles. As such the current acceptable limits of variations and a robust QA process as described in section 6.4 are currently deemed acceptable for these fault types. The focus of this research was therefore on the influence of the various carriageway fault types.

RCAs also use the rating data to monitor condition trends, usually by pavement use or hierarchy.

The influence of visual condition rating results and how they are used is described in following sections.

### 3.2.1 Surface Condition Index (SCI)

In this section we establish the impact of the visual rated parameters on the Surface Condition Index (SCI) value reported for the network. We look at establishing our level of confidence in the data as this performance measure can be used to track trends and make comparisons between different RCA networks.

The SCI value is calculated using:

$$SCI = \min(100, (CI + AI)) \quad \text{(Equation 3.1)}$$

Where:

$$CI = \min(100, \max(4 * ACA + 0.5 * ARV + 80 * APT + 20 * APH + 1.2 * AFL)) \quad \text{(Equation 3.2)}$$

$$AI = 3 * \min(100, \max(0, ((AGE2 - SLIF) / (SLIF * 12)))) \quad \text{(Equation 3.3)}$$

Where:

CI = Condition Index

AI = Age Index

ACA = percentage of alligator cracking

ARV = percentage area of ravelling

APT = percentage area of potholes

APH =percentage area of pothole patches.

AFL = percentage area of flushing

Where:

ACA = 0.0004 * sqr (alligator * 50 / insp_length) + (0.28 * alligator * 50 / insp_length)

ARV = 100 * scabbing / insp_area

APT = 100 * 0.05 * holes / insp_area

APH = 100 * 0.125 * patch / insp_area

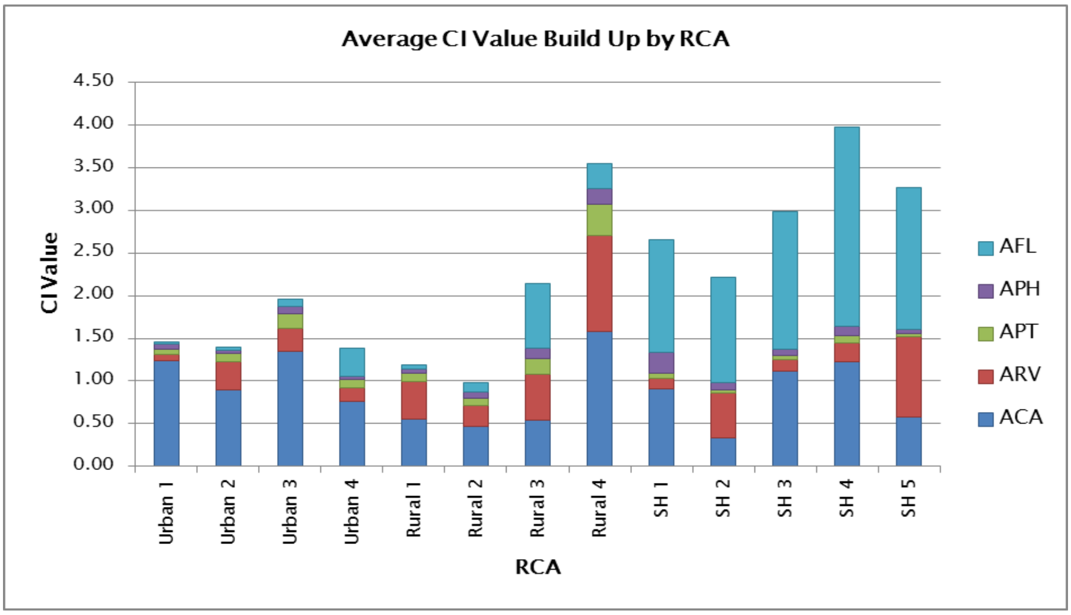
AFL = (flushing * 1.0 / insp_area) * 100

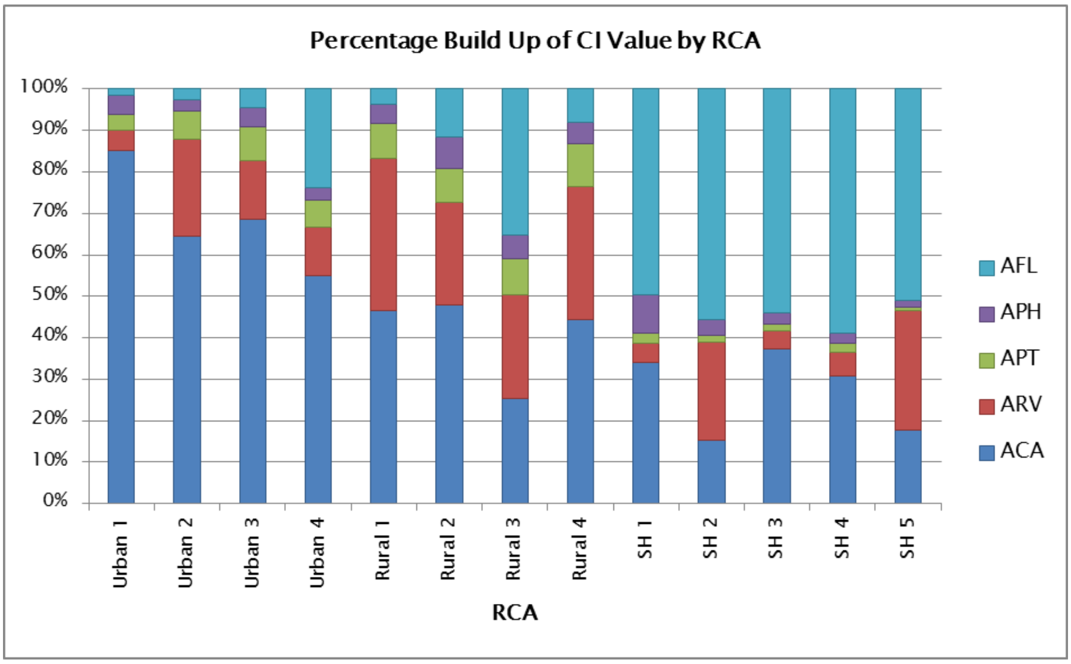
AGE2 = year (today) – year (surface_date

SLIF = expected surface life

A review of the data of four urban (urban 1, urban 2, urban 3 and urban 4), four rural (rural 1, rural 2, rural 3, rural 4 and rural 5) and five state highway (SH) networks (SH1, SH2, SH3, SH4 and SH5) has been undertaken to ascertain the sensitivity of the visual rating parameters on the overall network SCI value. Figure 3.1 shows the contribution of each parameter to the CI value.

**Figure 3.1 CI value build up for a cross-section of RCAs**





The charts in figure 3.1 show that alligator cracking and scabbing are the visual rating parameters with the greatest contribution to the CI value. Flushing, potholes and pothole patches also contribute but generally to a lesser extent.

Flushing is recorded by HSDC for the state highway network. There is a significant increase in the quantity of flushing recorded using this method. This is likely to be a result of these networks carrying higher traffic volumes which increase the presence of flushing. However, it should be noted that the HSDC uses a different mechanism to determine flushing compared with the visual condition rating. Therefore, there could be a step change in results from rating to HSDC.

A few trends can be seen from the data:

- Alligator cracking contributes a greater percentage of the CI value for urban networks.
- Scabbing contributes a greater percentage of the CI value on rural networks.
- Potholes make a smaller percentage contribution to the CI value on the state highway network.

It should be noted that there is a significant difference in typical SCI values for chip-sealed and asphalt-surfaced roads, due to the extent of cracking on asphalt (and slurry-sealed) surfaces. This issue is discussed in more detail in the NZTA research report (in progress) ‘Performance indicator analysis and applying levels of service’.

#### 3.2.1.1 Sensitivity check

To have confidence in the data used to calculate the SCI value we therefore need to understand the sensitivity each visual rating parameter has on the CI value. Our findings, based on a typical 2 x 3.5m lane, 50m long inspection length, are:

- Accurate alligator cracking data is important. Each 1m of alligator cracking recorded contributes 1.12 to the CI value for a 50m sample length. Therefore, if this value was 8, the current acceptable limits of variation would be 2 – 14. This gives a CI value range of 2.24 – 15.69. Consistent over or under-reporting could result in an inconsistent CI value.
- Potholes and pothole patches are also sensitive when calculating the CI value. A single pothole or patch contributes 1.14 and 0.71 respectively to the CI value for a 50m sample length. Therefore, for a value of 8 the acceptable range is 2 – 14 resulting in a CI value range of 2.29 – 16.00 for potholes and 1.43 – 10.00 for pothole patches. With such a large range, the confidence and consistency of the CI value for networks with a large number of these fault types reduces.
- Scabbing has the lowest input per percentage area into the overall CI value. For example, 10m2 of additional scabbing adds only 1.43 to the CI value. Over and under-reporting within the acceptable limits should therefore not drastically alter the CI value for the network.
- Each square metre of flushing adds 0.34 to the CI value. The current acceptable limits of variation are large for flushing as it is a category B fault type. Therefore a value of 8 has an acceptable range of 0 – 17. This equates to a range of 0 – 5.86 towards the CI value. Consistent over or under-reporting on those networks where flushing is common would reduce confidence in the data.

Based on the above, the following comments summarise each visual rating parameter used in calculating the CI value.

**Table 3.1 Effect of parameters on CI value**

| Parameter | Comments |
|---|---|
| Alligator cracking | This is a common occurring fault type (greatest on urban networks) which provides the greatest proportion of the CI value of visual rated parameters, particularly for urban networks.<br>The CI value would be affected by consistent over or under-reporting.<br>Confidence in the data for this parameter is therefore necessary to have an accurate and consistent CI value. |
| Scabbing/ravelling | This is the most commonly recorded fault type by inspection length.<br>It is the second highest contributor of the visual rated parameters to the CI value and contributes a similar percentage on rural networks to cracking.<br>On average, urban networks have a smaller contribution to the CI value.<br>Occasional over or under-reporting would have a negligible impact on the overall value.<br>Consistent over or under-reporting would reduce confidence and consistency in the data. |
| Potholes | Potholes are reasonably infrequent fault types, particularly on the state highway networks.<br>Potholes contribute only a small percentage of the CI value, <10%, and the values are very low for the state highway network.<br>The contribution to the CI value is, however, very sensitive to over or under-reporting; however; due to the infrequent nature of this fault type it is very unlikely to affect the CI value. |
| Pothole patches | This is an infrequent fault type contributing only a small amount to the CI value.<br>The contribution to the CI value is sensitive to over or under-reporting; however; due to the infrequent nature of this fault type any inconsistency in data is likely to have a negligible impact on the overall CI value. |
| Flushing | Flushing recorded as a visual rated fault generally makes only a small contribution to the CI value (with the exception of rural 3). However, the increase when captured by HSDC raises doubts over the quality of the data.<br>When recorded during the visual rating surveys this fault type is present on only 2% – 22% of inspection lengths.<br>Only consistent significant over or under-reporting would have any impact on the overall CI value. |

It should be noted that for SCI, shoving is not included in the index components. However, according to the visual condition rating manual, if other faults occurring within an area are affected by shoving, then only shoving is to be recorded. Therefore, areas of cracking that occur where shoving also occurs, will not be included in the SCI index. This may lead to some under-reporting of SCI values in areas with significant levels of shoving.

### 3.2.2 Pavement Integrity Index (PII)

This section establishes the impact of the visual rating parameters on the Pavement Integrity Index (PII) value. This value like the SCI value is reported at network level and can be used to track trends and compare between RCA networks. For this to be done we need confidence in the accuracy and consistency of the data.

The PII value is calculated using:

$$PII = \min (100, \max (0, rdm-8.4)*7 + \max (0, ACA - 3)*4 + \max (0, NAASRA - (\text{IF Urban } 120, \text{else } 90)) * 0.4 + ASH*3 + (APT+APH)*2) \qquad \text{(Equation 3.4)}$$

Where:

RDM = % of rutting. = hsd_rutting_avg OR ((0.366 * rutting / 2) * (50 / insp_length)) + 8.4
ACA = % of alligator cracking. = 0.0004 * sqr (alligator * 50 / insp_length) + (0.28 * alligator * 50 / insp_length.
NAASRA = average NAASRA
ASH = % of shoving. = 100 * shoving / insp_wheelpath
APT = % area of potholes = 100 * 0.05 * holes / insp_area
APH = % area of pothole patches. = 100 * 0.125 * patch / insp_area.

Therefore the 2 scenarios for IRI are:

- max(0,naasra_avg - 120) * 0.4 for urban roads
- max(0,naasra_avg - 90) * 0.4 for rural roads

It is important to note that the roughness portions of the equation, in particular, have threshold formats before roughness starts to contribute to PII. Therefore networks that have a number of sections with roughness levels around these thresholds are likely to have variations in their PII values with any changes in roughness.

A review of the data of four urban (urban 1, urban 2, urban 3 and urban 4), five rural (rural 1, rural 2, rural 3, rural 4 and rural 5) and five state highway networks (SH1, SH2, SH3, SH4 and SH5) has been undertaken to ascertain the sensitivity of the visual rating parameters on the overall PII value. Figure 3.2 shows the contribution for each parameter to the PII value. It excludes the roughness component as this is not a visual road condition rating parameter. The effect of the inclusion of roughness is shown in figure 3.3.

Figures 3.1 and 3.2 show that shoving and alligator cracking are the parameters with the greatest contribution to the PII value. Rutting also has occasional high contributions.

Rutting is recorded as part of the HSDC for the state highway networks. Four of the five state highway networks show large levels of rutting in comparison with the other networks. This is likely to be a result of these networks carrying higher traffic volumes, which increase the presence of this fault type. However, it should be noted that the HSDC uses a different mechanism to determine rutting compared with the visual condition rating. Therefore, there could be a significant change in results from using HSDC as a rating mechanism. Once the HSDC-surveyed average rut depth exceeds 8.4mm, the difference is multiplied by 7. This highlights the complexities created by having two such dissimilar methods in use.

An investigation was undertaken to determine the contribution visual rating parameters make to the overall PII value.

It can be seen that roughness generally contributes a greater percentage to the total PII value than all the visual rating parameters combined. As a result, any inconsistency in the visual rating data will reduce its impact on the PII value.

**Figure 3.2 PII value build up for a cross section of RCAs (roughness excluded)**

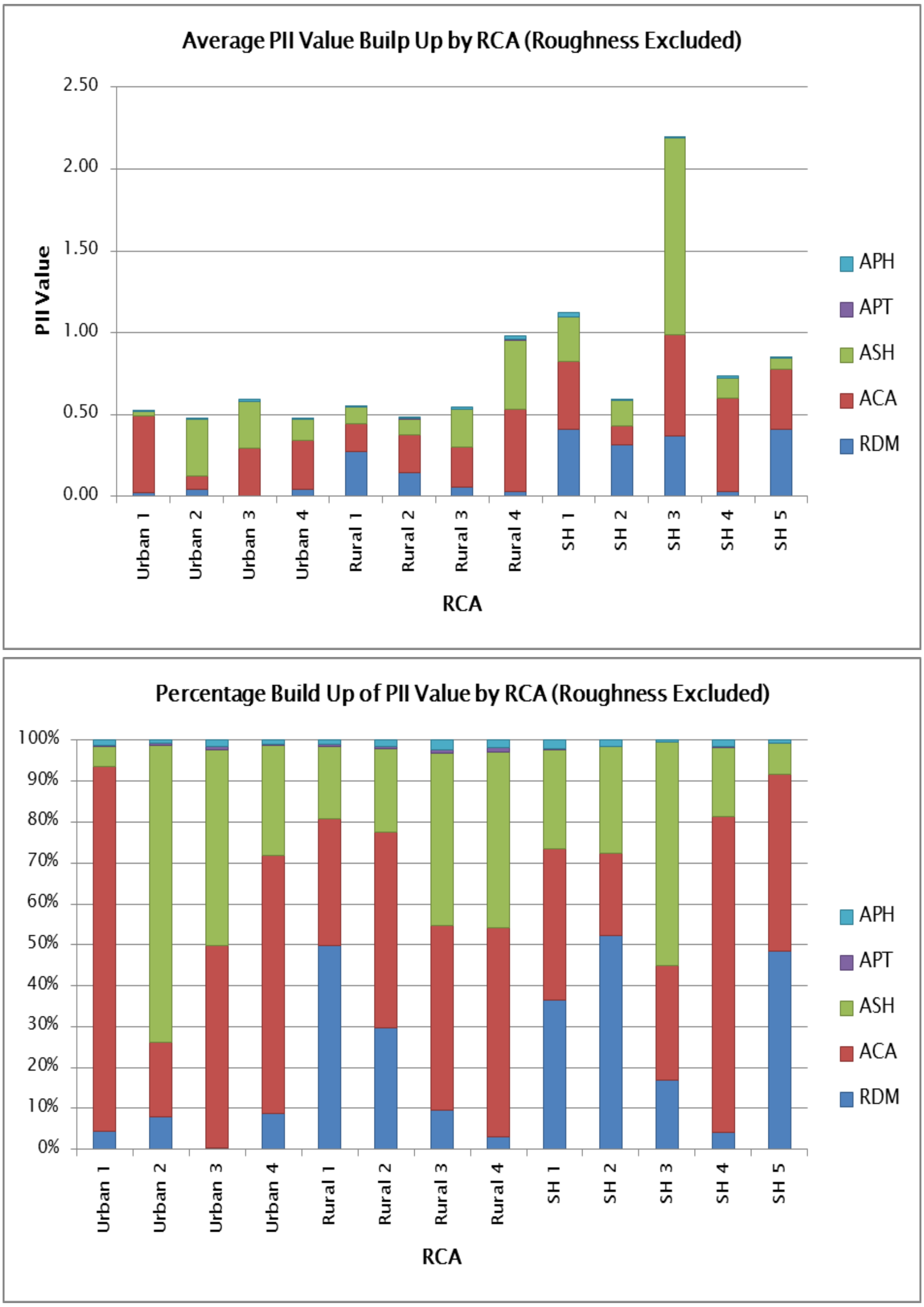
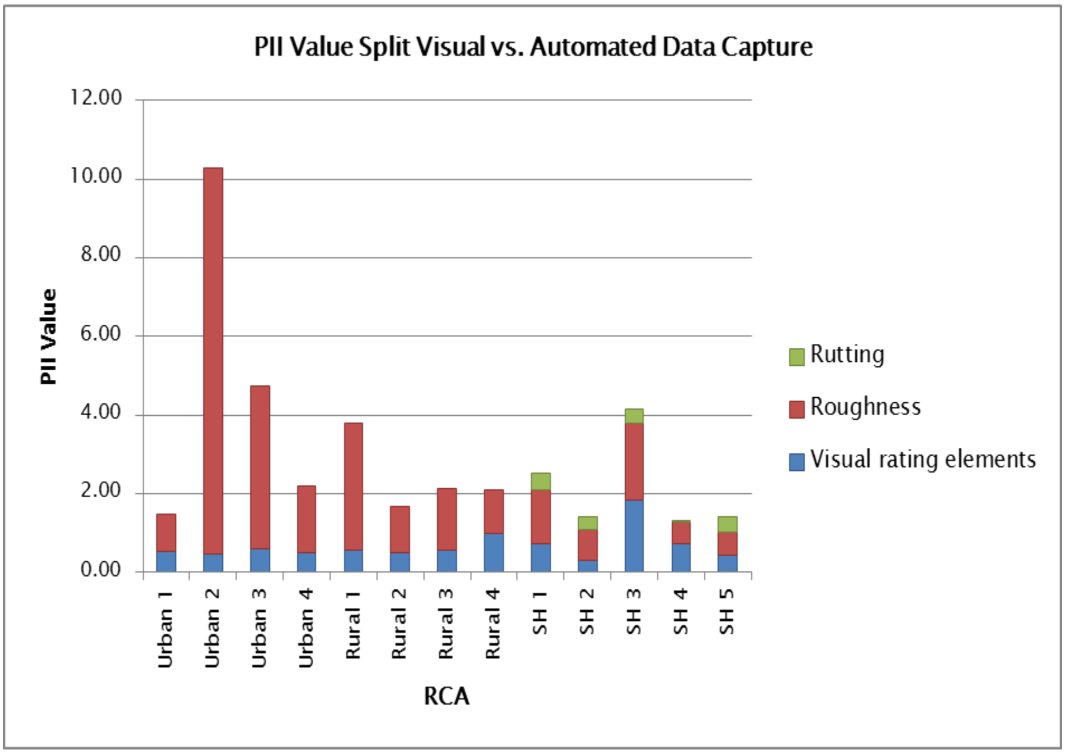
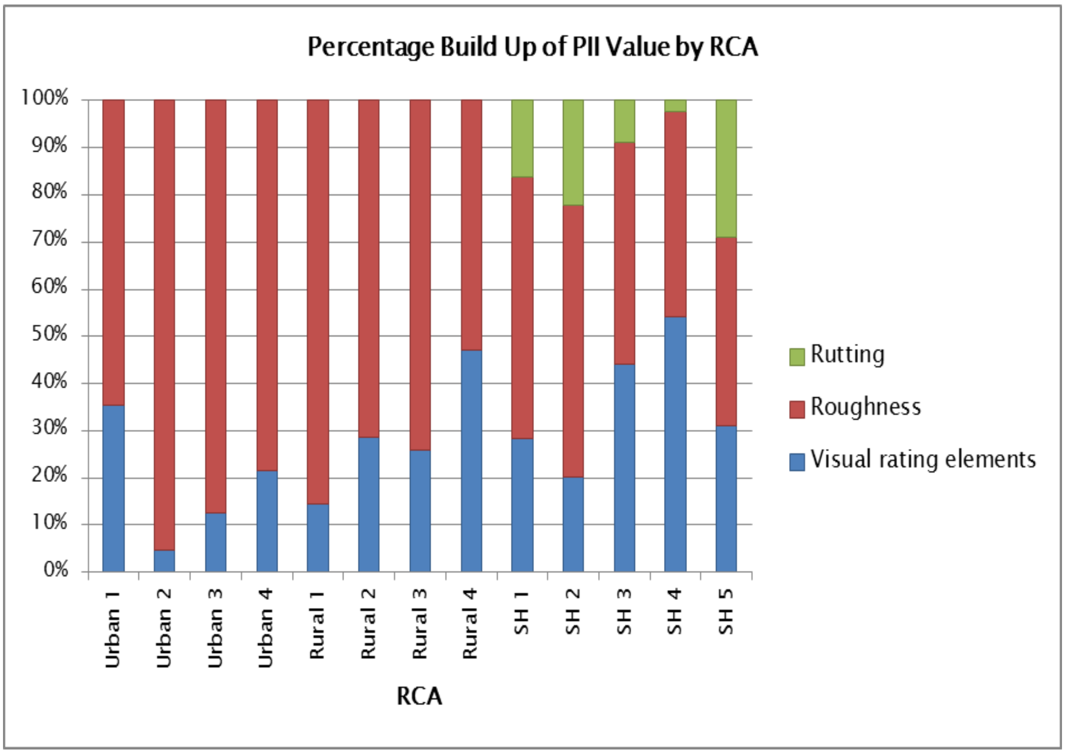


**Figure 3.3 PII value split visual vs automated surveys for a cross section of RCAs**





#### 3.2.2.1 Sensitivity check

To have confidence in the data used to calculate the PII value we therefore need to understand the sensitivity of each visual rating parameter. Our findings, based on a typical 2x3.5m lane, 50m long inspection length, are:

- Rutting is fairly sensitive to over or under-reporting. Each 1m of rutting recorded contributes 1.28 towards the PII value for that treatment length. Consistent over or under-reporting would be necessary to affect the overall PII value for the network with any significance.
- More than 10m (3%) of alligator cracking needs to be recorded in order to trigger input into the PII. Once the threshold is reached this parameter is reasonably sensitive to inaccurate data, with a 1m change altering the PII value for a treatment length by 1.12. Consistent over or under-reporting could have a small impact on the PII value.
- Shoving is very sensitive to over or under-reporting. Each 1m of shoving recorded contributes 1.50 to the PII value for that treatment length. Consistent over or under-reporting would have an impact on the overall PII value for the network.
- Potholes and pothole patches are not sensitive to over or under-reporting. Fairly large increases in the value recorded do not significantly affect the overall PII value.

Based on the above, the following comments summarise each visual rating parameter used in calculating the PII value:

**Table 3.2 Effect of parameters on PII value**

| Parameter | Comments |
|---|---|
| Rutting | This is a very low-frequency fault type which has a very low contribution to the PII value. The values are generally higher for the state highway networks where this parameter is recorded as part of the HSDC.<br>The parameter’s contribution to the PII value is sensitive to over or under-reporting. However, as it is an uncommon fault type on territorial local authority (TLA) networks, any inconsistency in data would have negligible impact on the overall value. |
| Alligator cracking | This is a reasonably frequent fault type and the greatest contributor to the PII of the visual rating parameters. It is more frequent on urban networks.<br>Due to the frequency and the contribution to the PII value being sensitive to over or under-reporting, any inconsistency will have a reasonable impact on the PII value. This impact is reduced on networks with greater roughness values. |
| Shoving | Shoving generally has a reasonable contribution towards the PII value.<br>This fault type occurs infrequently in the sample lengths.<br>The contribution towards the PII is very sensitive to over or under-reporting. Therefore, due to the uncommon nature of this fault type this would have reasonable impact on the overall PII value for the network, particularly those with lower roughness values. |
| Potholes | Potholes make a negligible contribution to the PII value.<br>They are an infrequent fault type and are not sensitive to over or under-reporting.<br>Any inconsistency in data would not affect the PII value. |
| Pothole patches | Pothole patches make a negligible contribution to the PII value.<br>They are an infrequent fault type and are not sensitive to over or under-reporting.<br>Any inconsistency in data would not affect the PII value. |
| Roughness | While roughness is not a visual condition rating parameter, it does play a significant role in the PII value. It averages 70% of the TLA PII values and 40% to 50% of the state highway network values. |
| Visual condition rating parameters | On TLA networks, the visual condition data makes up only 35% of the PII index. Therefore any influences from the visual condition data are limited. |

### 3.2.3 RAMM treatment selection algorithm (TSA) analysis

In this section we determine the impact of variations in the visual rating fault values recorded on the outputs from the RAMM TSA. We look at the trigger levels for treatments and analyse the sample data from the annual rating course to establish how the spread of values affects outputs from the TSA.

#### 3.2.3.1 Treatment selection procedure for thin-surfaced flexible pavements

The TSA runs a series of logical tests to identify any recommended treatment and provide a reason. The proposed treatments are based on certain trigger levels being met. Table 3.3 shows the levels at which visual rating parameters trigger a treatment for thin-surfaced flexible (TSF) pavements.

**Table 3.3 TSA trigger levels for TSF pavements**

| | **Treatment selected: Reseal in budget** | **Treatment selected: Reseal in next budget year** | **Treatment selected: Locking coat** |
|---|---|---|---|
| Alligator cracking | >3% wheelpath length | 1%–3% wheelpath length | - |
| Shoving | >3% wheelpath length | 1%–3% wheelpath length | - |
| Shoving + alligator cracking | >3% wheelpath length | 1%–3% wheelpath length | - |
| Potholes + pothole patches | >2.5% wheelpath length (>1 hole/patch per 20m of lane) | 2%–2.5% wheelpath length (1 hole/patch per 20m–25m of lane) | - |
| Scabbing | >25% carriageway area + >50% top surface life expectancy | 10%–25% carriageway area + >50% top surface life expectancy | >10% carriageway area + <50% top surface life expectancy |
| Flushing | >30% carriageway area | 15%–30% carriageway area | - |

#### 3.2.3.2 Sensitivity

Any proposed treatment is based on fault levels along a whole treatment length. Therefore the values are a sum of those for each inspection length within a treatment length.

Although most of the trigger levels are small it would take consistent over or under-reporting across all inspection lengths, or a single significant over or under-reporting on an inspection length, within a treatment length to have an impact on the proposed treatment.

Proposed treatments are validated through a visual site inspection by an experienced roading engineer.

#### 3.2.3.3 Spread of proposed treatments based on rating course data

During the 2011 annual visual rating condition course, a number of sample inspection lengths were rated by the course attendees.

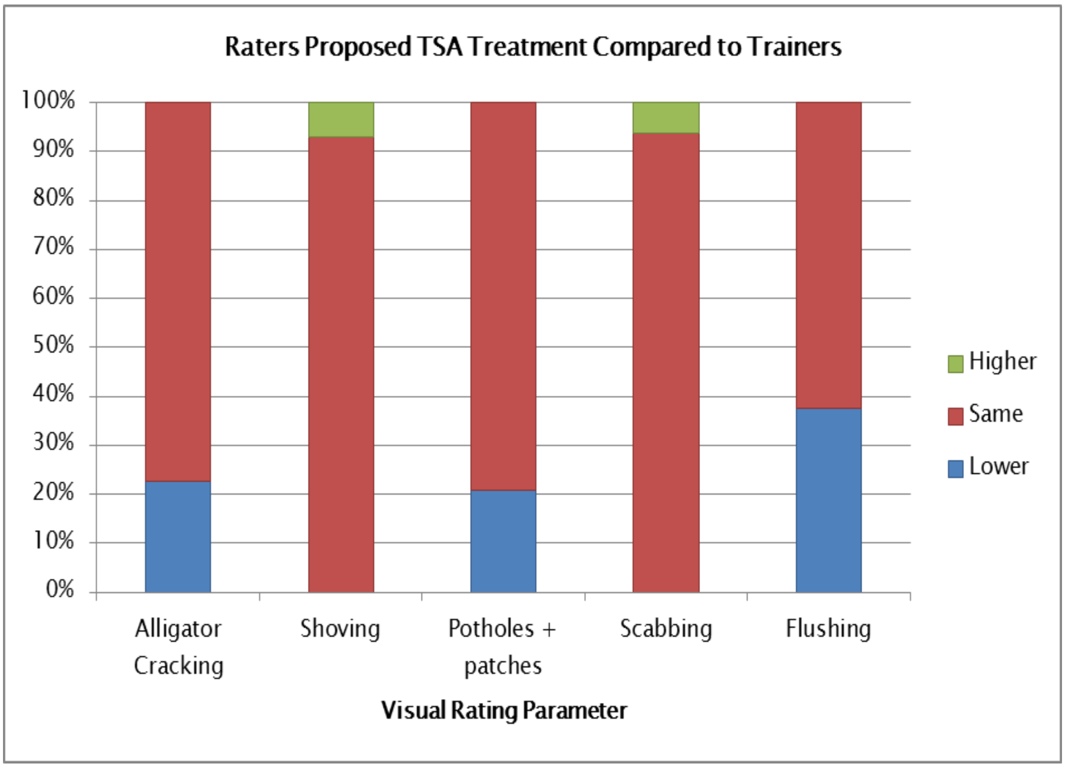
Figure 3.4 shows the proposed treatment variations for the sample lengths where the trainer recorded fault values around the trigger levels shown in table 3.3. Where the trainer's value is well in excess of the trigger values for 'reseal in budget' these have been ignored as the spread of results would not change the proposed treatment. For this exercise, locking coat treatments have not been separated and are reported as either 'reseal in budget' or 'reseal in next budget year'.

The inspection lengths for the course were 50m. The values triggering a treatment for this length are shown in table 3.4. The inspection lengths are assumed to be two lanes with a total width of 7m.

Table 3.4 **Treatment type trigger values for 50m inspection lengths**

| Pavement fault | Treatment selected: Reseal in budget | Treatment selected: Reseal in next budget year |
|---|---|---|
| Alligator cracking | >6m | 2-6m |
| Shoving | >6m | 2-6m |
| Shoving + alligator cracking | >6m | 2-6m |
| Potholes + pothole patches | >5 no. | 4-5 no. |
| Scabbing | >87.5m$^2$ | 35-87.5m$^2$ |
| Flushing | >115m$^2$ | 57.5-115m$^2$ |

Figure 3.4 **Comparison of the impact of raters' and trainer's values on TSA treatment**



It is evident that the raters' values for shoving and scabbing result in a treatment type consistent with the trainer's. There are a few instances of over-reporting resulting in a higher priority treatment being recommended. The visual site validation process should address these.

Alligator cracking, potholes plus pothole patches and flushing are of greater concern. Here we can see that between approximately 20% and 40% of the raters' values result in a lower priority treatment through the TSA triggers. The concern is if the error results in a change from 'reseal in next budget year' to 'general maintenance'. Those noted for 'general maintenance' would not necessarily receive the visual site validation to confirm the condition, thus preventing any correction. This would result in some lengths requiring treatment not being included in the FWP.

It was not possible to analyse any data for alligator cracking plus shoving as there were no inspection lengths containing both fault types at or around the trigger values.

Further investigation showed that 100% of the alligator cracking and flushing values recorded by the raters, which changed the TSA treatment, resulted in a change from ‘reseal in next budget year’ to ‘general maintenance’. Approximately 50% of the raters’ values of potholes plus pothole patches had the same outcome.

This is a concern as a large percentage of the treatment lengths requiring a ‘reseal in the next budget year’ are being missed from the FWP. The effect of this is likely to be increased maintenance costs and disruption to a smooth FWP following the next rating survey.

The definition and rating of potholes and pothole patches are well defined given the nature of the fault. However, cracking and flushing are of concern in their variability which affects the prediction of treatments and the selection of sites for inspection.

### 3.2.4 NZdTIMS modelling

In this section we establish how the visual rating parameters contribute to the outputs from the dTIMS pavement modelling programme. This programme is used to establish up to 20-year FWPs, therefore, confidence in the data is needed for producing a robust programme.

The identification of a treatment need in dTIMS is triggered typically by the SII value and other individual fault parameters exceeding a predetermined level in the trigger model. These levels are set by the RCAs and vary for treatment type, such as reseal, rehabilitation, etc as well as varying for road hierarchy, such as local, arterial, etc. In the optimised models, SII can be treated as a proxy for condition.

Some visual rating parameters along with the surfacing expected life make up the SII value.

$$SII = \min(100, (4*ACA + 0.5*ARV + 80*APT + 1.2*AFL + 3*\max(0, (AGE2 - SLIF) / SLIF * 12))) \qquad \text{(Equation 3.5)}$$

Where:

SII = Surface Integrity Index

ACA = area of cracking in % = 0.0004 * sqr (alligator * 50 / insp_length) + (0.28 * alligator * 50 / insp_length)

ARV = area of ravelling in % = 100 * scabbing / insp_area

APT = area of potholes in % = 100 * 0.05 * holes / insp_area

AFL = area of flushing in % = (flushing * 1.0 / insp_area) * 100

AGE2 = surface age in years = year (today) – year (surface_date)

SLIF = expected surface life in years = expected surface life

The contribution of the visual rating parameters to the SII value is the same as that for the SCI value but with the exclusion of pothole patches. The sensitivity of the visual condition parameters will therefore be the same as for the SCI value as pothole patches have only a minimal impact on SCI value.

Alligator cracking and potholes have the greatest sensitivity to under or over-reporting. Also relatively small values of those being recorded could trigger a treatment. Values greater than 2.5% alligator cracking or one pothole approximately every 11m contribute a factor greater than 10 to the SII value for that treatment length. This could be enough to trigger a treatment on certain road types for some RCAs. Again cracking is likely to be the most variable as the pothole fault is well defined and shows little variability in the surveys.

These are similar to the trigger levels within the TSA for a 'reseal in budget'. The sensitivity of these two parameters to over or under-reporting can be taken to be the same as for the TSA. Flushing and scabbing are much less sensitive with 8.5% and 20% respectively of the treatment area to be affected to contribute 10 to the SII value. Again, these are not too dissimilar to the trigger levels in TSA. Over or under-reporting of these parameters would need to be significant to affect the treatment selection from dTIMS.

### 3.2.5 Condition trends

Visual condition rating data is used by RCAs to monitor condition trends on their network. This is typically done by defect type over say the last 10 years split either by pavement use or hierarchy. To have confidence in these trends consistency in the data is needed between surveys.

This is particularly important for fault types exhibiting pavement and surfacing failure (alligator cracking, rutting, shoving and flushing).

### 3.2.6 Summary of the influence of rating parameters on condition indices

The results of the influence of the visual rating parameters on the performance measures are summarised in table 3.5.

**Table 3.5 Influence of visual condition rating parameters on performance measures**

| Parameter | TSA calculation | Condition trends | SCI | PII | dTIMS |
|---|---|---|---|---|---|
| Surface water channels | 3 | 3 | | | |
| Rutting | 3 | 1 | | 2 | 1 |
| Shoving | 1 | 1 | | 2 | 2 |
| Alligator cracking | 1 | 1 | 1 | 2 | 1 |
| Longitudinal and transverse cracking | 3 | 2 | | | 3 |
| Joint cracking | 3 | 2 | | | 3 |
| Potholes | 1 | 2 | 2 | 3 | 2 |
| Pothole patches | 1 | 2 | 2 | 3 | 2 |
| Edgebreak | 3 | 3 | | | 3 |
| Edgebreak patches | 3 | 3 | | | 3 |
| Scabbing (ravelling) | 2 | 2 | 1 | | 2 |
| Flushing | 2 | 1 | 1 | | 1 |

1= Core to process, 2=Moderate importance, 3=Used but not of significance

From this, those parameters with greatest influence are: rutting, shoving, alligator cracking, scabbing and flushing. These parameters will be investigated further to understand how the distribution of values impacts on the outputs for the performance measures.

The rating manual should be changed to include the recording of alligator cracking where it is located within an area of shoving. Although the alligator cracking is a secondary fault, as a result of the shoving it is core to a number of the performance measures listed in table 3.5. If this fault type is not reported, it will result in an under-reporting of the performance measures.

# 4 Limits of variation

## 4.1 What is the purpose of the limits of variation?

The limits of variation are set up to define the extent that any inspection value can deviate from the auditor's values before causing concern. The concerns are that:

- the inspected value is incorrect
- the rater is incorrectly rating faults and needs to be corrected
- the rater is consistently under or over-reporting fault quantities
- the level of accuracy is not sufficient for the purposes of the data
    - Pavement and Surfacing Condition Indices allow reliable analysis
    - dTIMS and TSA triggers are reliable

The limits of variation must meet all of these needs to work effectively.

## 4.2 What are the current limits of variation?

Within the PFM 6, faults recorded can have an acceptable limit of variation during QA checks and still be deemed acceptable. These limits of variation were established prior to the majority of the performance measures being established.

To have confidence in the performance measures we therefore need to ensure that the limits of variation are fit for the current and any potential future use of the visual rating data. Current limits of variation are split into three categories based on the level of variability allowed. Table 4.1 contains the visual rating parameters by category for sealed roads.

**Table 4.1 Visual rating parameters by category**

| Category A | Category B | Category C |
|---|---|---|
| Alligator cracking | Rutting | Inadequate drainage |
| Shoving | Flushing | Ineffective shoulder |
| Potholes | Scabbing | Blocked SWC |
| Pothole patches | Joint cracking | Inadequate SWC |
| | Longitudinal and transverse cracking | |
| | Edgebreak | |
| | Edgebreak patches | |
| | High lip | |
| | Broken surface | |
| | Blocked channel | |
| | Broken channel | |
| | Uphill channel | |

The parameters that have the greatest influence on the outputs for which they are used are alligator cracking, potholes, pothole patches, shoving, rutting, flushing and scabbing. Of these, the first four are category A, and the remainder category B.

The acceptable limits of variation for the three fault categories are shown below. These have been produced using the formula in section 3.5 of the PFM 6. The current acceptable limits of variation are:

Category A: $L = \pm 2 \times \sqrt{Va}$ where $Va > 12$ (Equation 4.1)

$L = \pm (\frac{1}{4} Va + 4)$ where $Va <= 12$ (Equation 4.2)

Category B: $L = \pm 3 \times \sqrt{Va}$ where $Va > 12$ (Equation 4.3)

$L = \pm (\frac{1}{4} Va + 7)$ where $Va <= 12$ (Equation 4.4)

Category C: $L = \pm 4 \times \sqrt{Va}$ where $Va > 12$ (Equation 4.5)

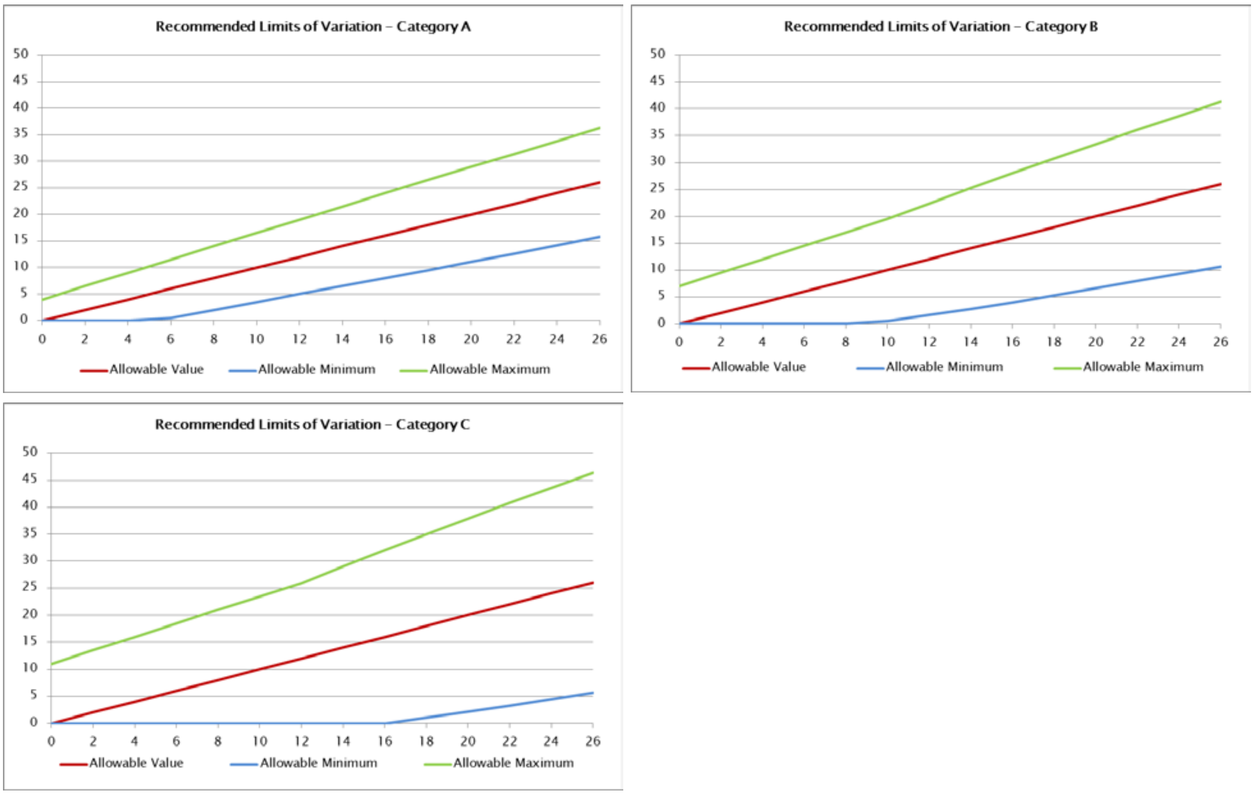
$L = \pm (\frac{1}{4} Va + 11)$ where $Va <= 12$ (Equation 4.6)

Where:

L = limit of variation

Va = value of defect measured by auditor

**Figure 4.1 Acceptable limits of variation for the three categories**



These charts show that the recommended acceptable limits of variation are tightest for category A, then B and finally C.

## 4.3 What level of consistency is achieved in the field?

We compared the variability of results from a number of raters who assessed the same sections of road during the 2011 annual rating certification course.

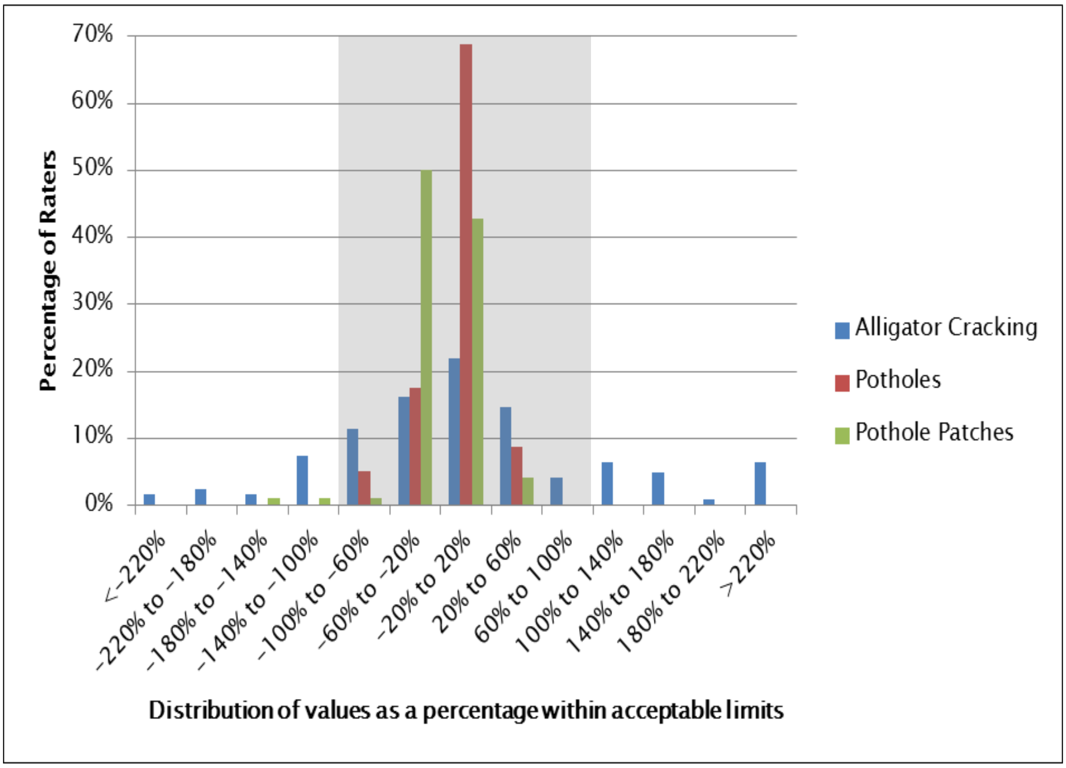
### 4.3.1 Review of the annual rating certification course February 2011 data

The February 2011 course results showed the raters' values were spread across the full acceptable limit of variation for all parameters. The limits of variation were not consistent for all parameters. The following sections analyse the data for the parameters that have most impact on the network performance outputs for which they are currently used. These are alligator cracking, potholes, pothole patches, shoving, rutting, flushing and scabbing.

#### 4.3.1.1 Alligator cracking

Results for alligator cracking are of concern with the distribution of values falling well outside the acceptable limits (shaded portion) as can be seen in figure 4.2.

**Figure 4.2 Consistency between trainer's and raters' values for alligator cracking, potholes and pothole patches**



The percentage variability of the raters can be seen in figure 4.2. The figure plots the percentage of raters on the training course assessing the fault quantity within the listed percentage value bands of the trainer's value. For alligator cracking, the raters' values are widely spread. This raises issues over confidence with the data for this parameter and is likely to affect consistency between different RCAs and between survey years. This will also impact on pavement and condition indices and their level of confidence.

A check on misidentification of the type of cracking suggested this was not the cause. A tightening of these limits would result in increased non-conformance.

#### 4.3.1.2 Potholes

Figure 4.2 shows a level of consistency between the raters’ and the trainer’s identification of pothole faults. This consistency should provide confidence in the data.

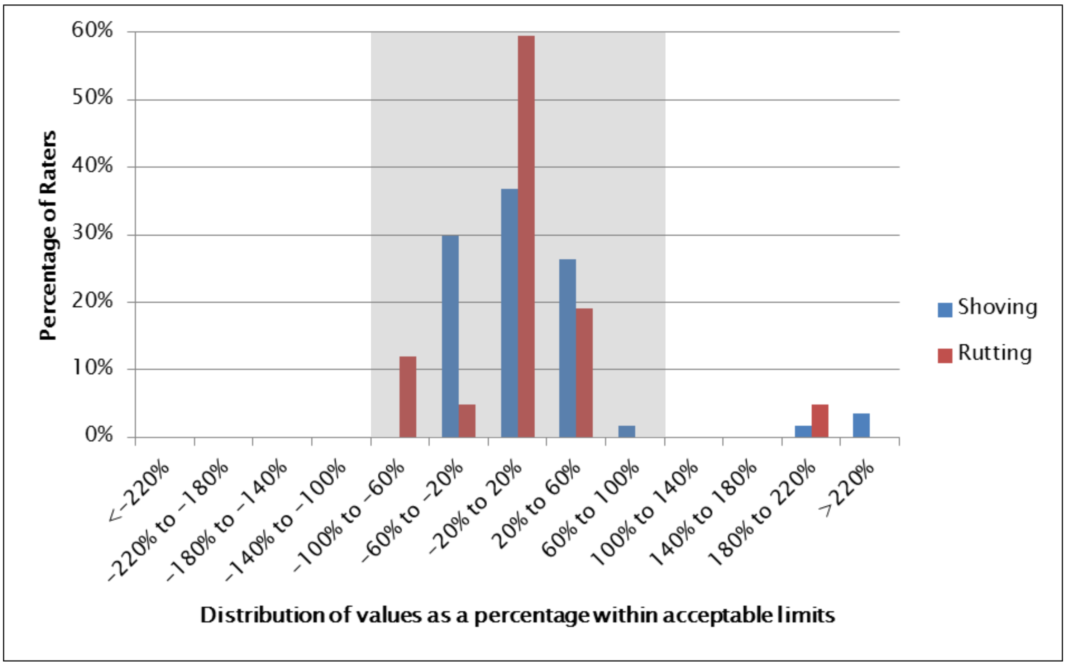
The current acceptable limits of variation could be tightened for this parameter without an increase in raters’ values falling outside the limits. Due to the raters’ values being small for this parameter one additional pothole recorded (or not) could result in up to a 100% variation. This suggests that changing the acceptable limits to a percentage from the ‘actual’ is not practical for small fault values.

The consistency of the raters’ values for this parameter is probably due to the low level of subjectivity involved in identification of potholes and the recorded value being a count of the number of instances.

#### 4.3.1.3 Pothole patches

The raters’ values for pothole patches are very similar to those for potholes, as would be expected, with regards to the distribution and consistency when compared with the trainer’s value. However, figure 4.2 shows a trend for raters to under-report. An understanding of the sensitivity of the under-reporting is needed to understand the impact of this inaccuracy. What are the intervention levels (triggers) at which a treatment is needed/not needed or an increase in treatment type (reseal to rehabilitation, etc) required? This may need to be investigated further but is not considered critical to the visual condition rating surveys.

**Figure 4.3 Consistency between trainer’s and raters’ values for shoving and rutting**



#### 4.3.1.4 Shoving

There were only four inspected sections in which shoving was recorded. The distribution of the raters’ values can be seen in figure 4.3. These figures show there is a reasonable level of consistency between the raters’ and trainer’s values. However, there are a number of raters’ values which fall well outside the current limits of acceptable variation.

Of the three raters' values above the acceptable limits of variation, one of them could be a misidentification and should have been recorded as rutting. An understanding of the sensitivity of under and over-reporting this parameter is needed to achieve improvements in consistency and confidence.

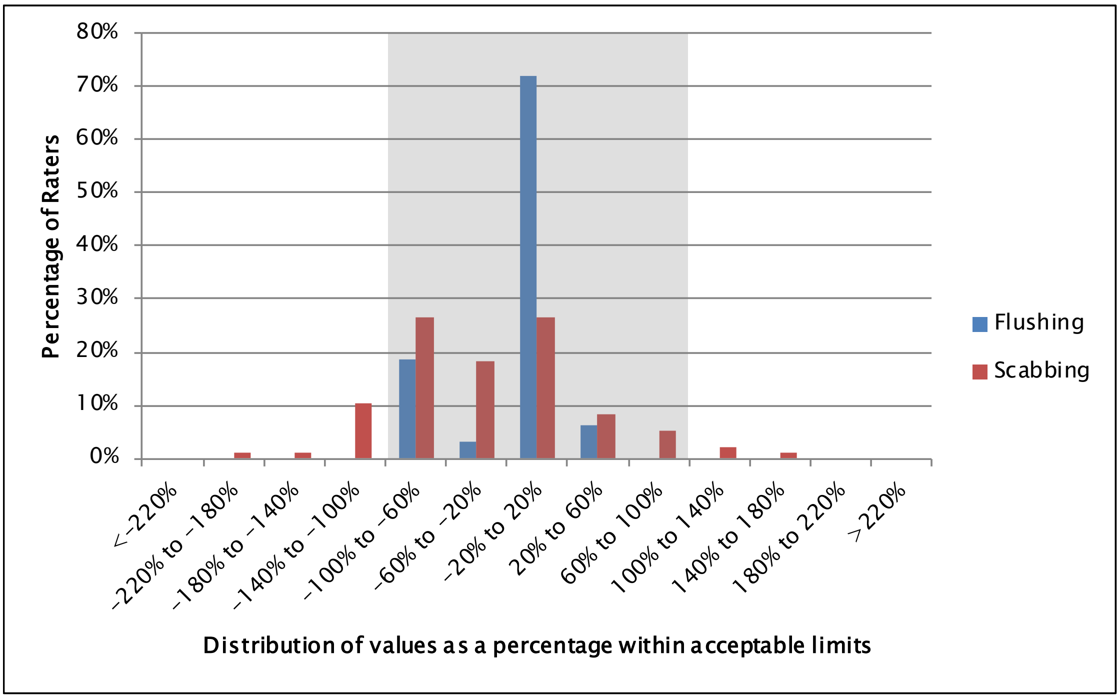
#### 4.3.1.5 Rutting

Again little data for rutting was captured during the annual certification course. There were only three inspection lengths where rutting was identified. Figure 4.3 shows reasonable consistency between the raters' and trainer's values. There was one exception where two raters measured rutting values of 15 when the trainer, and none of the other raters, identified any rutting.

#### 4.3.1.6 Flushing

As with shoving and rutting there is very little data from the annual courses for flushing. What data is available shows good consistency when compared with the trainer's values and a slight trend towards under-reporting as shown in figure 4.4.

**Figure 4.4 Consistency between trainer's and raters' values for flushing and scabbing**



Flushing is currently a category B fault. Consideration could be made to changing it to category A as a result of the increased impact it has on the outputs for which it is currently used. This re-categorisation would result in a small amount of raters' values from the annual certification course falling below the acceptable limit.

#### 4.3.1.7 Scabbing

The raters' values for scabbing show an increasing spread as the trainer's value increases. This is quite possibly down to the level of subjectivity associated with identifying this fault.

There are a high percentage of raters' values that fall outside the current acceptable limits. Therefore consistency and confidence in the data between RCAs and inspection years is low.

#### 4.3.1.8 Overview

Overall there was a good correlation between the trainer's and raters' values for the faults identified during the 2011 annual certification course as shown in figure 4.5. For the majority of the parameters analysed the recorded values generally fell within the current acceptable limits of variation. The distribution of values is such that a tightening of the limits for numerous parameters would not have significantly affected compliance.

The extent of the distribution of raters' values for alligator cracking and scabbing is of concern. The level of variation shown gives little confidence in the accuracy of the data and any consistency between RCAs or inspection years. A means of improving the identification of these parameters is necessary.

**Figure 4.5 Consistency between trainer's and raters' values overview and average**

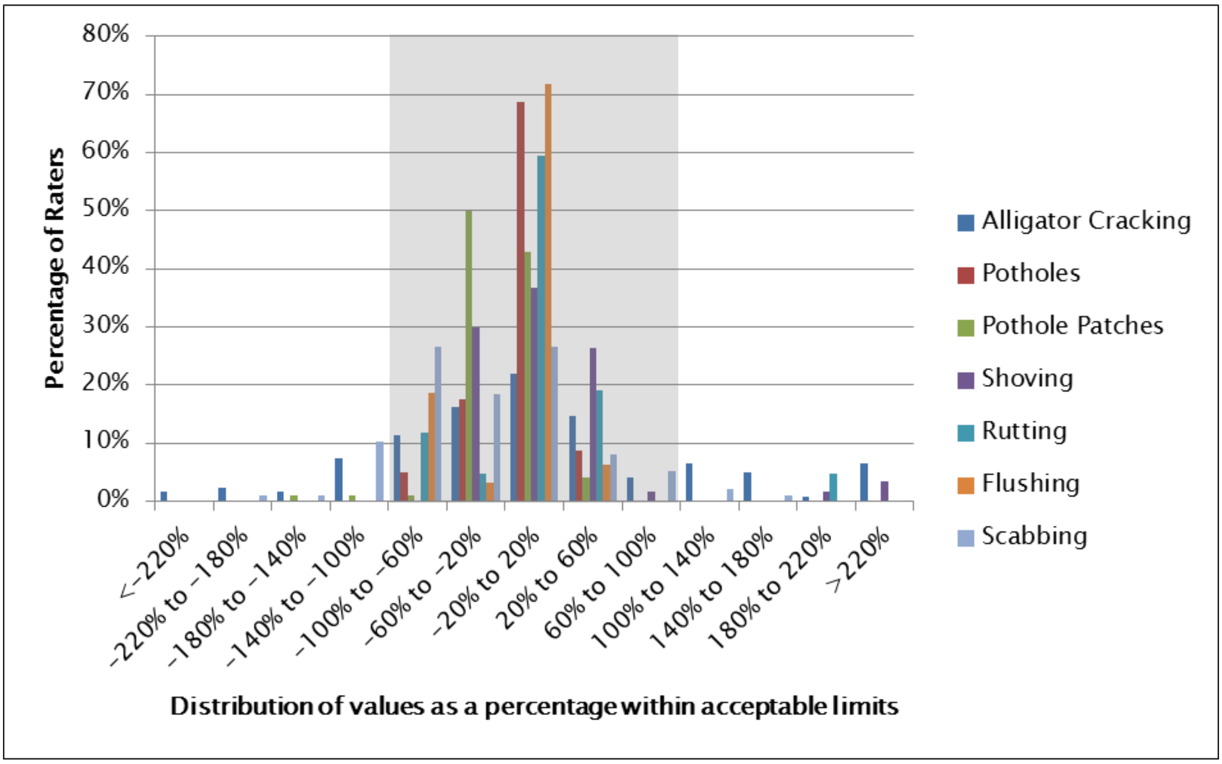


**Table 4.2 Variability of key visual rating parameters on performance measures**

| Parameter | TSA calculation | Condition trends | SCI | PII | dTIMS | Variability concern | Variability impact |
|---|---|---|---|---|---|---|---|
| Rutting | 3 | 1 |  | 2 | 1 | N | N |
| Shoving | 1 | 1 |  | 2 | 2 | N | N |
| Alligator cracking | 1 | 1 | 1 | 2 | 1 | Y | Y |
| Potholes | 1 | 2 | 2 | 3 | 2 | N | N |
| Pothole patches | 1 | 2 | 2 | 3 | 2 | N | N |
| Scabbing (ravelling) | 2 | 2 | 1 |  | 2 | Y | Y |
| Flushing | 2 | 1 | 1 |  | 1 | N | N |

1= core to process, 2=Moderate Importance, 3=used but not of significance

Table 4.2 summarises the findings regarding which parameters are variable and important enough to cause concern. Cracking and ravelling (scabbing) are two parameters whose variability is most likely to affect the condition index values and TSA/dTIMS treatment predictions. This impact is assessed more thoroughly in the following section.

# 4.4 Impacts from the current limits

## 4.4.1 Current acceptable range

The current acceptable limits of variation as shown in section 4.2 permit a reasonably large distribution of raters' values that are deemed compliant. These limits were established in 1997 prior to the majority of current uses of the performance measures. To understand the impact of the upper and lower limits on the performance measures we looked at the trainer's values recorded during the annual rating course and analysed the impact of the current limits on the SCI, PII, TSA and dTIMS. The results of this can be found in table 4.3 and show that for the SCI and PII the current acceptable limits of variation can result in a significant difference in the under or over-reporting value. This is particularly the case for alligator cracking. This parameter is used in the same way to calculate the SII value for treatment lengths as an input in dTIMS. The large variation is likely to result in a misidentification of treatment type.

Table 4.3 also shows how the current limits can affect the output from the TSA. Only 11 of 20 (55%) of the identified treatments were not altered when applying values at the upper and lower tolerance limits. These resulted from the rated value being substantially greater or less than the trigger levels, therefore resulting in either 'general maintenance' or 'reseal in budget'.

For those values rated around the trigger levels, applying the upper and lower tolerance limits results in a change in identified treatment. This is a concern as over or under-reporting within the current acceptable limits will reduce confidence in any short-term FWP produced.

**Table 4.3 Impact of current acceptable limits of variation on data uses**

| Fault by rated section | | | | | | % area | | | SCI/SII contribution | | | PII contribution | | | TSA identified treatment | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Lower limit | Trainer's value | Upper limit | Length (m) | Width (m) | Lower limit | Trainer | Upper limit | Lower limit | Trainer | Upper limit | Lower limit | Trainer | Upper limit | Lower limit | Trainer | Upper limit |
| Alligator crack PN1 | 22 | 33 | 44 | 50 | 7 | 6.02 | 9.24 | 12.46 | 24.10 | 36.97 | 49.84 | 12.10 | 24.97 | 37.84 | Reseal year 1 | Reseal year 1 | Reseal year 1 |
| Alligator crack PN2 | 17 | 27 | 37 | 50 | 7 | 4.65 | 7.56 | 10.47 | 18.61 | 30.25 | 41.89 | 6.61 | 18.25 | 29.89 | Reseal year 1 | Reseal year 1 | Reseal year 1 |
| Alligator crack PN5 | 0 | 5 | 10 | 50 | 7 | 0.00 | 1.40 | 2.87 | 0.00 | 5.60 | 11.49 | 0.00 | 0.00 | 0.00 | Gen maint | Reseal year 2 | Reseal year 1 |
| Alligator crack A1 | 15 | 25 | 35 | 50 | 7 | 4.20 | 7.00 | 9.80 | 16.81 | 28.01 | 39.21 | 4.81 | 16.01 | 27.21 | Reseal year 1 | Reseal year 1 | Reseal year 1 |
| Alligator crack A2 | 22 | 33 | 44 | 50 | 7 | 6.02 | 9.24 | 12.46 | 24.10 | 36.97 | 49.84 | 12.10 | 24.97 | 37.84 | Reseal year 1 | Reseal year 1 | Reseal year 1 |
| Alligator crack A3 | 27 | 40 | 53 | 50 | 7 | 7.66 | 11.20 | 14.74 | 30.64 | 44.81 | 58.98 | 18.64 | 32.81 | 46.98 | Reseal year 1 | Reseal year 1 | Reseal year 1 |
| Alligator crack A4 | 24 | 36 | 48 | 50 | 7 | 6.72 | 10.08 | 13.44 | 26.89 | 40.33 | 53.77 | 14.89 | 28.33 | 41.77 | Reseal year 1 | Reseal year 1 | Reseal year 1 |
| Alligator crack P1 | 0 | 2 | 7 | 50 | 7 | 0.00 | 0.56 | 1.82 | 0.00 | 2.24 | 7.28 | 0.00 | 0.00 | 0.00 | Gen maint | Reseal year 2 | Reseal year 1 |
| Alligator crack P4 | 13 | 23 | 33 | 50 | 7 | 3.76 | 6.44 | 9.13 | 15.02 | 25.77 | 36.51 | 3.02 | 13.77 | 24.51 | Reseal year 1 | Reseal year 1 | Reseal year 1 |
| Shoving PN3 | 0 | 1 | 5 | 50 | 7 | 0.00 | 0.50 | 2.63 | NA | NA | NA | 0.00 | 1.50 | 7.88 | Gen maint | Gen maint | Reseal year 2 |
| Shoving A1 | 0 | 0 | 4 | 50 | 7 | 0.00 | 0.00 | 2.00 | NA | NA | NA | 0.00 | 0.00 | 6.00 | Gen maint | Gen maint | Reseal year 2 |
| Shoving A2 | 0 | 1 | 5 | 50 | 7 | 0.00 | 0.50 | 2.63 | NA | NA | NA | 0.00 | 1.50 | 7.88 | Gen maint | Gen maint | Reseal year 2 |
| Shoving P4 | 0 | 0 | 4 | 50 | 7 | 0.00 | 0.00 | 2.00 | NA | NA | NA | 0.00 | 0.00 | 6.00 | Gen maint | Gen maint | Reseal year 2 |
| Scabbing PN4 | 0 | 9 | 18 | 50 | 7 | 0.00 | 2.57 | 5.21 | 0.00 | 1.29 | 2.61 | NA | NA | NA | Gen maint | Gen maint | Gen maint |
| Scabbing PN5 | 5 | 18 | 31 | 50 | 7 | 1.51 | 5.14 | 8.78 | 0.75 | 2.57 | 4.39 | NA | NA | NA | Gen maint | Gen maint | Gen maint |
| Scabbing A1 | 14 | 30 | 46 | 50 | 7 | 3.88 | 8.57 | 13.27 | 1.94 | 4.29 | 6.63 | NA | NA | NA | Gen maint | Gen maint | Reseal year 2 |
| Scabbing A2 | 14 | 30 | 46 | 50 | 7 | 3.88 | 8.57 | 13.27 | 1.94 | 4.29 | 6.63 | NA | NA | NA | Gen maint | Gen maint | Reseal year 2 |
| Scabbing P1 | 1 | 10 | 20 | 50 | 7 | 0.14 | 2.86 | 5.57 | 0.07 | 1.43 | 2.79 | NA | NA | NA | Gen maint | Gen maint | Gen maint |
| Scabbing P2 | 0 | 0 | 7 | 50 | 7 | 0.00 | 0.00 | 2.00 | 0.00 | 0.00 | 1.00 | NA | NA | NA | Gen maint | Gen maint | Gen maint |
| Scabbing P5 | 10 | 25 | 40 | 50 | 7 | 2.86 | 7.14 | 11.43 | 1.43 | 3.57 | 5.71 | NA | NA | NA | Gen maint | Gen maint | Reseal year 2 |

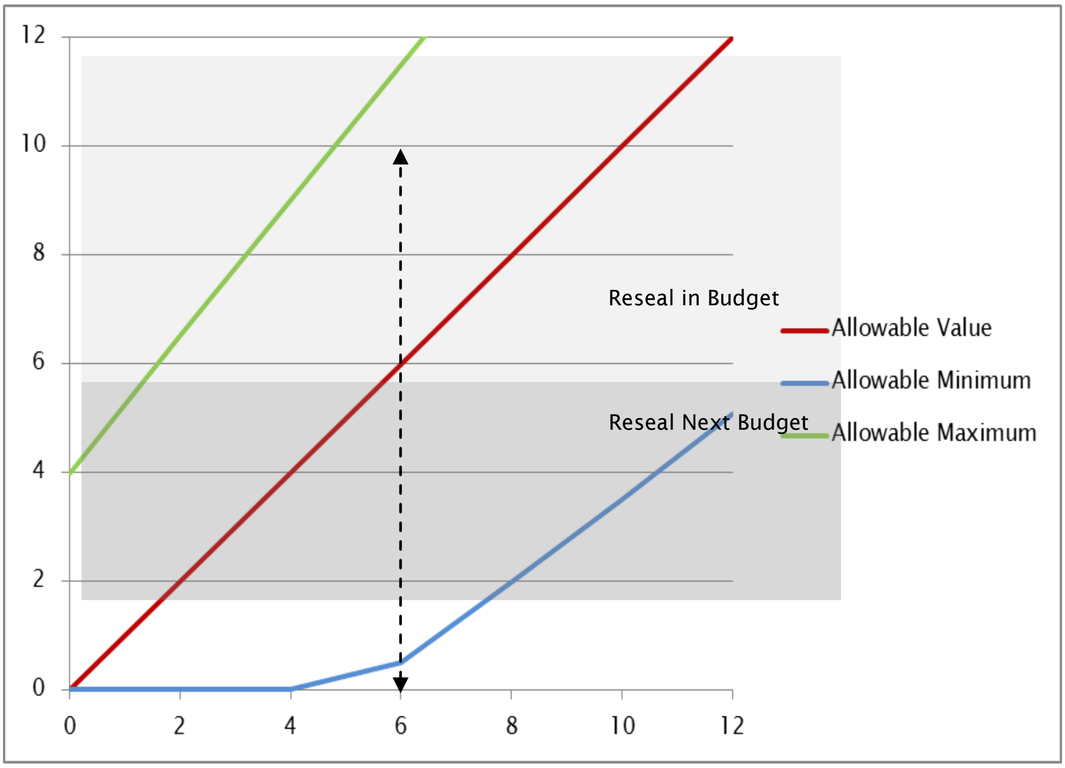
### 4.4.2 Small fault values

For small fault values the allowable variation is proportionally large. For example, alligator cracking has an acceptable variation of 60% or greater for values of 10m or less. As can be seen from the sample RCA data, fault values for all visual rating parameters recorded are generally small. This is also the case for RCAs that test 100% sample lengths which tend to be longer than 10m. Therefore, the greatest quantity of data recorded still has the largest allowable percentage variation falling within the acceptable limits of variation.

It is at the smaller fault value levels that treatments are triggered in TSA and dTIMS. Figure 4.6 shows the fault values at which the TSA triggers a treatment for alligator cracking for a typical 50m, 2 x 3.5m lane inspection length with the x axis representing the actual values. This figure shows that only a small variation in the recorded fault value is needed to change the treatment identified. For example, for an actual value of 5 the acceptable limits of variation allow the rated value recorded to be such that it can trigger anything from no treatment (general maintenance) to 'reseal in budget year'. We recommend tightening the limits around typical trigger level fault values to avoid this variation in outcome.

The use of smaller, more frequent inspection lengths will improve accuracy and increase the correct triggering of TSA treatments compared with the actual condition.

**Figure 4.6 Impact of current acceptable limits of variation on TSA treatment identified**



It should be noted that currently only 10% of rural and state highway network inspection lengths report the presence of cracking. So for a network of approximately 2500 inspection lengths, a QA process is performed on 5%, or 125 forms. Of these, only about 15 would show the presence of cracking. For a team of two raters, eight forms each would indicate cracking. From this sample, we would hope to determine whether there is a pattern of under or over-reporting. However, often only one form falls outside the limits of variation so it is unlikely the current method will identify any of the issues we are looking to mitigate through the QA process, see section 4.1. We therefore need to amend the whole QA process, not just the limits of variation.

### 4.4.3 Impact of acceptable limits of parameters on other reporting outputs

As detailed in section 3.2 the following parameters have been found to affect the result of the outputs for which they are now used:

**Table 4.4 Effect of visual rating parameters on performance measures**

| Performance measure | Effect of visual rating parameters |
|---|---|
| SCI | Alligator cracking is the largest contributor of the parameters to the CI value. Potholes, flushing and scabbing also make a contribution but to a lesser extent. |
| PII | Shoving and alligator cracking are the visual rating parameters that make any real contribution to the PII value. However, on a network with a high roughness value the impact of the visual rating parameters is significantly reduced as roughness will make up the majority of the PII value. |
| TSA | Low visual rating parameters can affect the output from TSA with only a small variation in the value. Parameters that are generally under-reported in value are of greatest concern. These have been seen to be alligator cracking, flushing and potholes plus patches. Under-reporting may result in a change from 'reseal in next year budget' to 'general maintenance' where there is no site validation. |
| dTIMS | Like TSA, a small variation in the rated value can affect the output from the model for small fault values, particularly alligator cracking and potholes. This could produce a FWP which undergoes reasonable alterations following the next rating round. This reduces confidence in the output from the model. The identified treatments from dTIMS are validated through a field inspection prior to producing any FWP. These field inspections are carried out by an experienced pavement engineer and provide a reasonable QA process to avoid identifying sections for treatment that is not required. |

## 4.5 How effective are the limits of variability?

The annual condition rating course data shows raters predominantly record fault values within the acceptable limits of variations for most parameter types, with the exception of alligator cracking and scabbing.

For smaller fault values, <=12, the raters were predominantly within the limits. The acceptable limits appear therefore to be generous for small fault values. For example, in table 4.3, alligator cracking from site PN5 show all raters' values within the required limits of variation. However the distribution of values would provide TSA results from 'general maintenance' to 'reseal in budget'.

The distribution of raters' values for alligator cracking and scabbing repeatedly fell outside the acceptable limits particularly for values >12. The distribution of values showed a large spread of under and over-reporting for these two parameters. This was probably a result of the subjective interpretation and/or difficulty in identifying these parameters. The analysis of the various RCAs' data shows these two fault types to be the most frequently occurring on their networks. A need to improve the identification of these is therefore important. However, the median fault value for both these parameters based on the RCAs' survey data is <6, whereas the annual certification course raters' values all show better correlation with the acceptable limits.

However, for the typical 50m per 500m inspection length commonly undertaken in the industry these fault value sizes are around the trigger levels in TSA for a treatment. The current limits of variation are wide enough to permit rated values resulting in anything from ‘general maintenance’ to ‘reseal in budget’ from TSA and can reasonably impact the SCI or PII values currently used as a measure of network condition. As a result the limits should be tightened to avoid a high level of under or over-reporting misrepresenting the actual condition.

### 4.5.1 Sensitivity of limits

As can be seen in chapter 3, variation in values within the current acceptable limits of variation can affect the result of the outputs for which they are now used. Typical fault values are small as shown in the RCA data in section 5.3. It is at these small values that the current acceptable limits of variation give the largest percentage error.

The SCI and PII values are reported at a network level. The visual rating data would therefore need to be consistently inaccurate in one direction to affect the value. Also, in the case of the PII value, the roughness data captured by HSDC is the largest proportion of the value. This reduces any impact of inaccuracy in the visual rating data.

However, the TSA is very sensitive to variations in the data when small fault values are involved. A small change can result in a different treatment being proposed. The TSA reports results per treatment length; therefore those treatment lengths consisting of multiple sample lengths are less likely to misidentify the treatment for isolated inaccuracies in the fault value recorded. Also, any proposed treatment is validated during a site visit by an experienced roading engineer.

With regards to the TSA it is those treatment lengths where fault values have been under reported, so that only general maintenance is proposed in the FWP, where a reseal is potentially the necessary treatment. This could lead to a higher level of maintenance cost than anticipated as well as a knock on effect on the development of a smooth FWP.

dTIMS is most sensitive to variations in alligator cracking and potholes values recorded, especially with smaller values at which the variations are likely to trigger the need for treatment. Like the TSA, dTIMS identifies any need for treatment based on treatment length. Therefore, those treatment lengths consisting of multiple sample lengths would need to be consistently under or over reported, or one significant misreport, to affect the outcome.

### 4.5.2 Proposed improvement areas and cost implications

Based on analysis of the data from the annual certification course and various RCAs, we recommend the following improvements:

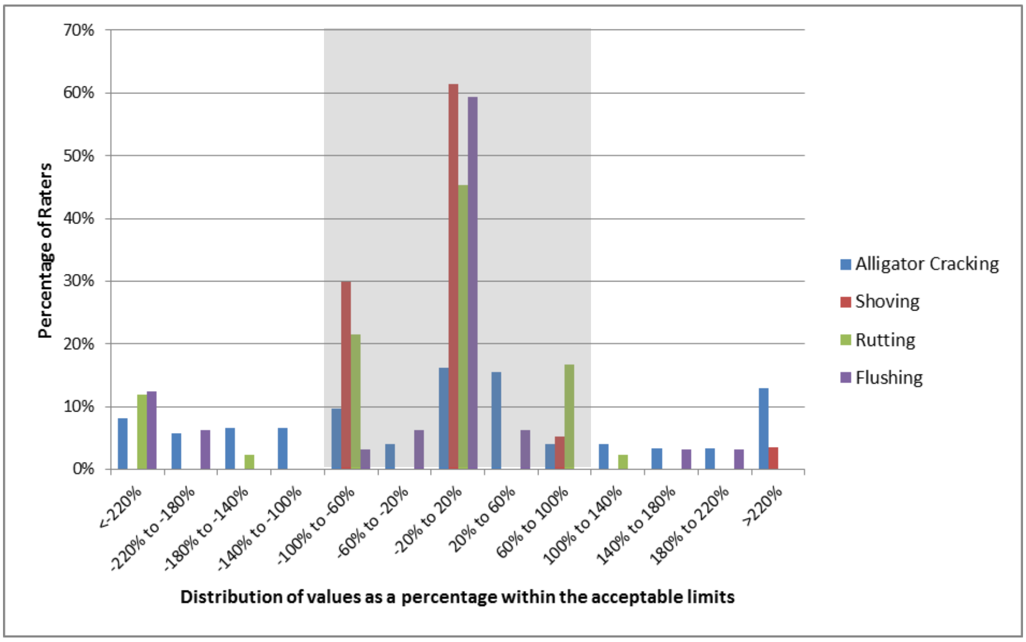
- As highlighted in the review of RCA data, targeting the accuracy of the smaller fault values should provide most benefit for confidence in the data. The smaller faults are also more likely to trigger treatments in the TSA and dTIMS. This paired with figure 4.7 shows the existing limits appear generous for smaller fault values. It should be possible to reduce these limits with only negligible/ marginal expected survey cost increases.
- Establishing a new category for the fault types with a high influence on the outputs for which they are used (alligator cracking, shoving, rutting and flushing). The category A limits of variation should be much tighter than they are at present. Proposed revised limits are:

$$L = \pm 1.2 \times \sqrt{Va} \qquad \text{where } Va > 12 \qquad \text{(Equation 4.7)}$$

$$L = \pm (\tfrac{1}{4} Va + 1) \qquad \text{where } Va <= 12 \qquad \text{(Equation 4.8)}$$

Figure 4.7 shows the compliance of the raters during the annual rating course with these tighter limits. The shaded area represents the limits of variation. It shows that, with the exception of alligator cracking, there would be very little increase in the number of non-compliances with the limits. This would help reduce the impact of under-reporting associated with the TSA and dTIMS.

**Figure 4.7 Distribution of raters' values against the trainer's – shown with tighter limits of variation**



- The current category B faults (rutting, flushing and scabbing) that have a significant influence on their outputs, should be changed to category A. We anticipate this will result in increased survey time and cost but will also produce a greater level of confidence and consistency in the data.
- Rutting and flushing data can be collected by HSDC to reduce the subjective nature. This will result in increased survey costs but the speed, confidence and consistency of the data captured will be of benefit. The annual condition rating course showed a trend of under-reporting flushing which had an impact on the treatment needs when run through TSA.
- Change the sample regime. For a 10% sample percentage a 20m per 200m gives a better indication of condition than the current 50m per 500m. Increasing the minimum sample percentage/length would provide data that better represents the condition of the treatment length. An additional cost would be associated with this but consistency and increased confidence in the data should be recognised. This is discussed in more detail in section 6.1.

## 4.6 What is best way of quantifying defects?

A change from the current method of collection requires justification on three fronts:

- The current method is providing results that are too variable.
- The impact of that variability affects the purposes for which the data is used.
- The method does not provide data in a form that is appropriate for its intended use.

]

From this research only cracking and scabbing fall into the first two categories. Rutting qualifies under the third.

The alternative would be a condition and extent type rating system. The condition type assessment would not have the definitive nature of the current measurement system. (However, given the current level of variability, would the loss of accuracy be critical if a measure of repeatability could be created?) A change of measuring system would also create issues with defining trigger criteria for TSA and dTIMS treatments. In addition the condition indices would need to be adjusted to factor in a subjective value.

This can be investigated further but it is important to trial other measures first, such as improved QA procedures to provide more consistent results.

### 4.6.1 Rutting

Rutting is increasingly used as a measure of pavement performance. Currently it is rated by reporting the length of a wheelpath rut depth greater than 30mm. However high-speed data surveys report a measure of average rut depth in each wheelpath. This form is much more useful for modelling purposes and the continuous data stream allows greater statistical analysis of the distribution of rut depth.

Ideally, the visual rating and HSDC survey methods would deliver measures that were in compatible or comparable units, even if there were differences in data accuracy.

It should be noted that the PFM 6, section 3 details two alternative methods for assessing rutting visually. The first is as described above whereby a length of rutting exceeding 30mm is reported. The specifications require that this is done by checking the depth of rutting under a 2m straight edge. However, this is usually gauged with the eye. This first method is used by most local authorities, certainly by all we are aware of. For the second method, 10 measurements are made in the outside wheelpath in each direction, at the start and end of the inspection length and at quarter points in between. The PFM 6 again specifies that the measurements are taken using a 2m straight edge. The RAMM data entry process allows the entry of the 10 readings into the mean rut depth field as a list separated by plus signs (+). The RAMM software then calculates the mean and standard deviation for the data entered.

There are therefore a series of options for improving the assessment of rutting:

- Option 1: Maintain the status quo of reporting the wheelpath length that exceeds 30mm, as currently done by RCAs.
- Option 2: Utilise the second method as detailed in the PFM 6. This creates data compatible with the high-speed data surveys and requires no updates to the RAMM software. However, the requirement for a measurement using the 2m straight edge means that raters would have to physically enter the traffic lane. While this would be mitigated by the presence of a spotter, there would still be a safety risk. The use of two-person teams would also greatly increase the cost of the surveys which are currently undertaken by a single person on Code of practice for temporary traffic management (COPTTM) designated low-volume and level 1 type roads.
- Option 3: Utilise the second method but allow visual assessment to the nearest 5mm (eg 0–5mm, 5–10mm, 10–15mm, 15–20mm, etc bands). This would allow a single person to continue doing the rating and again require no update to the RAMM software. There would be a significant drop in data confidence, but it would still be an improvement in usefulness compared with the first method.
- Option 4: Measure using high-speed data.

Adoption of option 3 would provide an assessment in line with HSDC methodology and a form of measure useful for condition indices, dTIMS modelling and assessments of pavement health. Although we anticipate it would slow the visual rating process and result in higher costs, it would still be much cheaper than the more definitive measurement of using the straight edge, requiring two-person teams. Raters would need to be well trained and have sound QA procedures put in place to monitor consistency.

It would also give each RCA the option of mixing manual visual rating and HSDC to measure rutting depending on their network needs and budget, while providing data in a format that would be comparable across the network. It would also improve the usefulness of trend reporting for this fault type. Showing the change and distribution of rut depth across the network is much more useful than tracking the change in length exceeding a 30 mm rut depth.

It is understood that option 3 does have drawbacks concerning relative consistency. However, it is important that this be viewed against the disadvantages of the current system, given the increased focus on network benchmarking, modelling and trend analyses.

## 4.7 Discussion regarding fault collection

### 4.7.1 Are parameters correctly categorised?

Of the parameters with most influence on the outputs for which they are used, alligator cracking, shoving, potholes and pothole patches are category A faults with the tightest limits of acceptable variation. However, rutting, flushing and scabbing are category B. The acceptable variation limits for category B faults are wider which could reduce confidence and consistency in the data. As this data is being used for outputs beyond what was originally intended by the PFM 6, consideration should be given to reclassify rutting, flushing and scabbing as category A. Alternatively a tighter limit could be introduced but this may be overkill for the other faults currently listed as category B. This will again result in increased inspection cost but should provide greater confidence in the data outputs.

### 4.7.2 Method of survey

A level of subjectivity is used when identifying the extent of a number of the fault types, particularly alligator cracking, scabbing and flushing. Any level of subjectivity is likely to reduce the consistency in the data when comparing between different networks, or even between different raters in different years on the same network. To improve confidence in the data any subjective identification needs to be removed.

This could be done through using an alternative method of data capture, ie HSDC for flushing. This would result in increased inspection cost but would improve data consistency. Where the fault type cannot be captured by alternative methods the examples in the PFM 6 could be reviewed, and these specific parameters discussed in greater detail at the annual road rating certification course.

# 5 Review of RCAs' visual collection data

## 5.1 Overview

We undertook a review of sample data from 13 RCAs, four urban (urban 1, urban 2, urban 3 and urban 4), four rural (rural 1, rural 2, rural 3 and rural 4) and five state highway (SH1, SH2, SH3, SH4 and SH5) networks, for the parameters (alligator cracking, potholes, pothole patches, rutting, shoving, flushing and scabbing) with most influence on their use in forward work programming and performance measures.

Tables 5.1 and 5.2 contain the results of this review.

## 5.2 Current inspection regimes

The data shows that rural and state highway networks are rated with a lower percentage sample size than urban networks. Rural networks are typically rated on a 50m per 500m pattern. The state highway networks are also generally rated with a minimum 50m sample length but the percentage is typically higher as a result of shorter treatment lengths.

The urban networks have greatest variation in the percentage of samples rated. Of the four networks, two rated 100% of the length, the others varied based on hierarchy. The percentage rated is likely to be linked to budget restrictions. In theory those networks that have 100% rating coverage should produce data that best reflects their actual condition.

The short inspection lengths are a function of short treatment lengths. These short treatment lengths should be reviewed but we recommend adopting a minimum inspection length of 20m.

## 5.3 Typical fault values

The median fault values are small for the parameters reviewed. The median value recorded of those rated lengths with faults was generally $\leq 6$, with the exception of rutting in rural 1 and rural 3 and alligator cracking and scabbing on SH5 which are 10, 9, 10 and 15 respectively. The majority of values recorded therefore have the largest percentage margin of error still falling within the acceptable limits of variation. Targeting improving the accuracy for smaller fault values ($Va \leq 12$) would have most benefit on improving confidence and consistency in the data. As seen in the review of the annual course data the distribution of the raters' values is such that a tightening of the acceptable limits should not have a significant impact on the number of non-conformances but would improve the quality of the data.

Large fault values, however, have been recorded, particularly for the urban networks. These networks generally have a higher percentage of samples, which should provide data that better reflects the condition of the treatment length. On examining the largest values recorded, it appears that for a number of the fault types these are possible errors or misidentifications. For example, urban 2 recorded 180m of shoving on a 266m sample length and urban 3 recorded 220 potholes on a 250m sample length. Both of these appear high for the associated sample lengths. Because of the likely magnitude of these fault values, a reasonably larger variation would not affect the outputs. They are well above the 'trigger' values.

The frequency of faults is generally small for each parameter on rural networks, usually <10% with the exception of scabbing and a few other isolated values.

Viewing all the data, we see there is generally a low percentage of small fault values on each RCA's network.

Rural 4 has a higher frequency of fault occurrence in comparison with the other rural networks. This could be the result of the network being in a poorer condition (there may be more sections with defects but the percentiles are not higher) or the inspection team over-reporting fault values.

### 5.3.1 Alligator cracking

Alligator cracking is most common on urban networks. The median values recorded range from 1 to 10. For the rural and state highway networks these fault values are around the trigger levels for treatment as measured by the RAMM TSA for typical rating lengths.

### 5.3.2 Potholes and pothole patches

Potholes and pothole patches are infrequently recorded during the RAMM rating surveys. They are again most frequent on urban networks, particularly those rated as 100%. The median values are low ranging from 0 to 3.

### 5.3.3 Shoving

Shoving is another infrequent fault type except on the SH3 network where shoving has been recorded on 17% of inspection lengths. The median value for all networks is small ranging from 1 to 3. Based on the typical rated lengths these are around the trigger levels for a treatment according to the RAMM TSA for rural and state highway networks.

### 5.3.4 Rutting

Rutting is the most infrequent fault type. It was recorded on ≤2% of inspection lengths. The median value is generally ≤3 with the exception of rural 1 and rural 3 which are 10 and 9 respectively. However this has much to do with the methodology whereby only lengths with a rut depth greater than 30mm are recorded. The importance of rutting as an indicator of pavement integrity requires a measure more frequently reported and one that better aligns with the HSDC surveys of continuous measurement of rut depth.

Rutting is measured by HSDC on the state highway network and therefore has not been reviewed.

### 5.3.5 Flushing

The frequency of flushing being recorded is quite inconsistent with the percentage of inspection lengths where flushing is identified as ranging between 2% and 22%. This does not appear to be related to the network being urban or rural. The median values are low ranging from 1 to 3.

Flushing is measured by HSDC on the state highway network and therefore has not been reviewed.

### 5.3.6 Scabbing

Scabbing is the most frequent parameter recorded for all the RCA networks with between 9% and 68% of inspection lengths exhibiting scabbing. The extent of scabbing varies widely with values ranging from 0m to 600m but the median values are small for each network, ≤5m$^{2,}$ with the exception of SH5, which is 15m$^2$.

**Table 5.1 Comparison of RCA survey data – category A visual rating parameters normalised to 50m inspection lengths**

| | | Sample length (m) | | | Sample % | | | Alligator cracking | | | | | Potholes | | | | | Pothole patches | | | | | Shoving | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RCA | Forms | Min | Max | Median | Min | Max | Median | No. of forms with faults | % Forms with faults | 15th percentile | 85th percentile | Median with faults | No. of forms with faults | % Forms with faults | 15th percentile | 85th percentile | Median with faults | No. of forms with faults | % Forms with faults | 15th percentile | 85th percentile | Median with faults | No. of forms with faults | % Forms with faults | 15th percentile | 85th percentile | Median with faults |
| Urban 1 | 3620 | 1 | 64 | 40 | 5% | 100% | 26% | 524 | 14% | 1.3 | 16.3 | 4 | 78 | 2% | 1.2 | 2.5 | 1 | 144 | 4% | 1.3 | 2.5 | 1 | 26 | 1% | 1.3 | 3.6 | 1 |
| Urban 2 | 4497 | 8 | 792 | 109 | 100% | 100% | 100% | 1415 | 31% | 0.1 | 73.7 | 1 | 567 | 13% | 0.1 | 30.8 | 0.4 | 309 | 7% | 0.1 | 23.3 | 0 | 237 | 5% | 0.1 | 71.4 | 2 |
| Urban 3 | 4120 | 4 | 300 | 173 | 100% | 100% | 100% | 1574 | 38% | 0.4 | 5.6 | 1 | 590 | 14% | 0.2 | 1.3 | 0.4 | 468 | 11% | 0.2 | 1.9 | 1 | 296 | 7% | 0.3 | 4.5 | 1 |
| Urban 4 | 3265 | 3 | 799 | 50 | 5% | 100% | 66% | 603 | 18% | 0.2 | 6.7 | 1 | 341 | 10% | 0.1 | 1.4 | 0.4 | 226 | 7% | 0.1 | 1.4 | 0 | 147 | 5% | 0.2 | 2.9 | 1 |
| Rural 1 | 2165 | 2 | 82 | 50 | 7% | 100% | 10% | 188 | 9% | 1.0 | 11.0 | 4 | 80 | 4% | 1.2 | 2.5 | 1 | 84 | 4% | 1.0 | 3.9 | 1 | 22 | 1% | 1.0 | 12.6 | 3 |
| Rural 2 | 2972 | 1 | 300 | 42 | 4% | 100% | 10% | 120 | 4% | 1.5 | 26.7 | 6 | 78 | 3% | 0.1 | 30.8 | 2 | 80 | 3% | 1.0 | 10.2 | 3 | 50 | 2% | 1.0 | 5.0 | 2 |
| Rural 3 | 4335 | 1 | 920 | 50 | 5% | 100% | 10% | 223 | 5% | 2.0 | 21.7 | 5 | 271 | 6% | 0.2 | 1.3 | 2 | 214 | 5% | 1.0 | 7.5 | 3 | 67 | 2% | 1.0 | 21.2 | 5 |
| Rural 4 | 1485 | 6 | 170 | 50 | 9% | 71% | 10% | 303 | 20% | 1.0 | 11.5 | 3 | 202 | 14% | 0.1 | 1.4 | 2 | 107 | 7% | 1.0 | 7.0 | 2 | 88 | 6% | 1.0 | 5.8 | 3 |
| SH1 | 1741 | 4 | 80 | 50 | 10% | 100% | 18% | 211 | 12% | 1.0 | 15.5 | 3 | 58 | 3% | 1.0 | 2.9 | 1 | 153 | 9% | 1.0 | 3.9 | 2 | 91 | 5% | 1.0 | 6.0 | 2 |
| SH2 | 1427 | 5 | 80 | 50 | 10% | 100% | 14% | 119 | 8% | 1.0 | 25.8 | 5 | 32 | 2% | 1.0 | 3.1 | 1 | 90 | 6% | 1.0 | 10.2 | 1 | 51 | 4% | 1.3 | 9.2 | 3 |
| SH3 | 2094 | 1 | 80 | 50 | 5% | 100% | 11% | 234 | 11% | 1.3 | 24.1 | 5 | 62 | 3% | 1.0 | 5.0 | 2 | 99 | 5% | 1.0 | 7.5 | 2 | 364 | 17% | 1.0 | 7.0 | 3 |
| SH4 | 2348 | 1 | 80 | 50 | 5% | 100% | 14% | 297 | 13% | 1.3 | 28.9 | 5 | 103 | 4% | 1.0 | 3.0 | 1 | 145 | 6% | 1.0 | 7.0 | 1 | 51 | 2% | 1.0 | 4.5 | 2 |
| SH5 | 1298 | 1 | 80 | 50 | 5% | 100% | 10% | 48 | 4% | 1.0 | 39.9 | 10 | 25 | 2% | 1.0 | 2.0 | 1 | 65 | 5% | 1.0 | 3.9 | 1 | 8 | 1% | 1.1 | 10.0 | 3 |

**Table 5.2 Comparison of RCA survey data – category B visual rating parameters normalised to 50m inspection lengths**

| | | Sample length (m) | | | Sample % | | | Rutting | | | | | Flushing | | | | | Scabbing | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RCA | Forms | Min | Max | Median | Min | Max | Median | No. of forms with faults | % Forms with faults | 15th percentile | 85th percentile | Median with faults | No. of forms with faults | % Forms with faults | 15th percentile | 85th percentile | Median with faults | No. of forms with faults | % Forms with faults | 15th percentile | 85th percentile | Median with faults |
| Urban 1 | 3620 | 1 | 64 | 40 | 5% | 100% | 26% | 12 | 0% | 1.3 | 5.5 | 3 | 66 | 2% | 1.3 | 3.7 | 1 | 313 | 9% | 1.3 | 7.1 | 3 |
| Urban 2 | 4497 | 8 | 792 | 109 | 100% | 100% | 100% | 96 | 2% | 0.2 | 1.6 | 1 | 188 | 4% | 0.1 | 41.0 | 1 | 1519 | 34% | 0.5 | 10.0 | 2 |
| Urban 3 | 4120 | 4 | 300 | 173 | 100% | 100% | 100% | 5 | 0% | 0.3 | 0.5 | 0 | 428 | 10% | 0.3 | 3.6 | 1 | 1713 | 42% | 0.5 | 7.4 | 2 |
| Urban 4 | 3265 | 3 | 799 | 50 | 5% | 100% | 66% | 36 | 1% | 0.2 | 4.8 | 1 | 475 | 15% | 0.3 | 8.4 | 2 | 826 | 25% | 0.3 | 6.9 | 1 |
| Rural 1 | 2165 | 2 | 82 | 50 | 7% | 100% | 10% | 22 | 1% | 1.3 | 5.5 | 10 | 66 | 3% | 1.0 | 7.0 | 3 | 681 | 31% | 1.0 | 17.5 | 3 |
| Rural 2 | 2972 | 1 | 300 | 42 | 4% | 100% | 10% | 75 | 3% | 0.2 | 1.6 | 2 | 216 | 7% | 1.0 | 10.0 | 3 | 1033 | 35% | 1.0 | 8.9 | 3 |
| Rural 3 | 4335 | 1 | 920 | 50 | 5% | 100% | 10% | 16 | 0% | 0.3 | 0.5 | 9 | 717 | 17% | 2.0 | 50.0 | 5 | 826 | 19% | 2.0 | 25.0 | 5 |
| Rural 4 | 1485 | 6 | 170 | 50 | 9% | 71% | 10% | 9 | 1% | 0.2 | 4.8 | 3 | 329 | 22% | 1.0 | 5.0 | 2 | 1013 | 68% | 2.0 | 19.7 | 5 |
| SH1 | 1741 | 4 | 80 | 50 | 10% | 100% | 18% | Measured by HSDC | | | | | Measured by HSDC | | | | | 148 | 9% | 2.0 | 10.0 | 5 |
| SH2 | 1427 | 5 | 80 | 50 | 10% | 100% | 14% | | | | | | | | | | | 501 | 35% | 1.0 | 15.0 | 3 |
| SH3 | 2094 | 1 | 80 | 50 | 5% | 100% | 11% | | | | | | | | | | | 871 | 42% | 1.0 | 3.0 | 1 |
| SH4 | 2348 | 1 | 80 | 50 | 5% | 100% | 14% | | | | | | | | | | | 330 | 14% | 1.0 | 15.4 | 3 |
| SH5 | 1298 | 1 | 80 | 50 | 5% | 100% | 10% | | | | | | | | | | | 365 | 28% | 5.0 | 40.9 | 15 |

# 6 Data collection and analysis

## 6.1 How does the sampling regime affect results?

### 6.1.1 Analysis of sampling regime effects

Continuous sampling data such as HSDC and the LTPP sites can be used to assess the impacts of different sampling regimes on the effectiveness of representing the condition of the road section for each fault.

A study Beca (1997) undertook for Transfund NZ produced the following correlation results:

**Table 6.1 Correlation coefficient of sample regime and continuous data (Beca 1997)**

| 10% sample size | | | 20% sample size | | |
|---|---|---|---|---|---|
| **Sample format** | **Correlation coefficient ($r^2$ value)** | **Regression formula** | **Sample format** | **Correlation coefficient ($r^2$ value)** | **Regression formula** |
| 50m per 500m | 0.45 | y=0.781x | 100m per 500m | 0.62 | y=0.902x |
| | | | 50m per 250m | 0.68 | y=0.838x |
| 20m per 200m | 0.76 | y=0.917x | 40m per 200m | 0.89 | y=0.977x |
| 10m per 100m | 0.82 | y=0.817x | 20m per 100m | 0.93 | y=0.823x |
| 10m per 100m* | 0.85 | y=1.123x | | | |

* 10m per 100m but offset 50m.

The regression formula gives the relationship between the actual results and the estimated results for each sample format. The actual result equals the ‘x’ value with the predicted result equalling the ‘y’ value.

The study, using a small sample size, showed that 10% sampling for 500m length gave a poor correlation of only 0.45 while 10% sample size taken every 100m showed much greater correlation with a coefficient of 0.85. Similarly, an increase in sample size from 10% to 20% increased the correlation. For example the 40m rating per 200m showed 0.89 correlation, and 20m per 100m showed 0.93 correlation.

For this current study, data from high-speed rutting and shoving surveys on eight networks (a 50/50 split of state highway and non-state highway), was extracted from the RAMM database for a total length of 2 x 10km per network. These were divided into 1km sections to simulate treatment lengths giving a total of 160 treatment lengths. The data was then analysed using a series of sample sizes and lengths factoring the values up to the 1km lengths. A 50m per 500m sampling format is common in the visual condition rating industry. However as the available high-speed data is only recorded in 20m lengths we looked at 40m per 500m and 60m per 500m which represent an 8% and 12% sample size respectively.

The sample lengths and frequencies used and the associated results are shown in tables 6.2 and 6.3. From this we can establish which of the sampling regimes has the best correlation with the actual data collected during the HSDC survey.

**Table 6.2 Correlation coefficient of sample regime and actual data**

| 10% sample size | | 20% sample size | | 40% sample size | |
|---|---|---|---|---|---|
| Sample format | Correlation coefficient ($r^2$ value) | Sample format | Correlation coefficient ($r^2$ value) | Sample format | Correlation coefficient ($r^2$ value) |
| 100m per 1000m | 0.02 | 200m per 1000m | 0.11 | 400m per 1000m | 0.52 |
| 40m per 500m | 0.30 | 100m per 500m | 0.60 | 200m per 500m | 0.72 |
| 60m per 500m | 0.40 | | | | |
| 20m per 200m | 0.76 | 40m per 200m | 0.77 | 80m per 200m | 0.83 |
| | | 20m per 100m | 0.85 | 40m per 100m | 0.85 |

Table 6.2 shows that the correlation increases, as expected, as both the sample size and frequency of samples increase. As sample size increases across the table, correlation increases significantly for infrequent sampling but much less significantly for more frequent sampling. Similarly correlation increases moving down the table. The improvement is much more marked in the 10% sample size than for the 40% sample size.

Table 6.2 suggests that a regime of 20m per 200m provides the most efficient solution. There is a strong correlation for a 10% sample rate with little further efficiency gained by increasing the percentage sample size. However, the Beca (1997) study showed a higher improvement in correlation with an increased sample size, 0.89 for 20%, see table 6.1.

A change from 50m per 500m to a minimum sample regime of 20m per 200m is expected to result in an increased survey cost caused by a slight reduction in rater productivity as the raters would be inspecting shorter inspection lengths on a more frequent basis. The reduced productivity is only anticipated to be applicable to rural networks due to their longer treatment lengths, and then to subsequent sample lengths associated with urban networks. This is supported by the typical sample lengths for the urban networks in tables 5.1 and 5.2. There will also be a negligible impact where the surface water channel is rated concurrently by walking the full length of the rating length.

**Table 6.3 Regression formula of sample format and actual data**

| 10% sample size | | 20% sample size | | 40% sample size | |
|---|---|---|---|---|---|
| Sample format | Regression formula | Sample format | Regression formula | Sample format | Regression formula |
| 100m per 1000m | y=0.3664x | 200m per 1000m | y=0.5053x | 400m per 1000m | y=0.7372x |
| 40m per 500m | y=0.5106x | 100m per 500m | y=0.9443x | 200m per 500m | y=1.0151x |
| 60m per 500m | y=0.8345x | | | | |
| 20m per 200m | y=1.1028x | 40m per 200m | y=1.0064x | 80m per 200m | y=0.9486x |
| | | 20m per 100m | y=0.9720x | 40m per 100m | y=0.8678x |

The regression formula gives the relationship between the actual results (x) and the estimated results (y) for each of the different sample regimes. Table 6.3 shows the results for the networks analysed. For a 10% sample size the actual and estimated values get closer as frequency increases. For 20% and 40% sample sizes there is no significant change for a rating length of 500m or less.

One note of concern is that as sample size and sample frequency decrease, the regression coefficient decreases. Ideally, one would prefer the regression coefficient to remain at around 1 with the scatter increasing as the confidence decreases. However, the change in the coefficient indicates that as the sample size and frequency decrease, the quantum of fault reported decreases. The implication is that for say SII, the result would be lower for a 50m per 500m sampling regime than for a 40m per 200m sampling regime by as much as 30%. Again, this trend is evident in the Beca (1997) report, although to a lesser extent.

This means that the consistency of the sampling regime is very important if the NZTA wishes to use rating data in condition indices across networks with different sampling regimes. The second implication is that there will likely be a step change in reported condition if RCAs change from one sampling regime to another. The NZTA should be aware of this in implementing any rating regime changes. An example is network urban 5 where a change from a 10% to a 100% sample size resulted in the CI value dropping from 0.98 to 0.52.

### 6.1.2 Impact of sampling regime effects

To understand the impact of the above findings we need to ascertain the typical treatment lengths for the various network types. This is summarised in table 6.4.

**Table 6.4 Summary of treatment lengths by network type**

| Network type | Network | Treatment length (m) | | | Rating regime |
|---|---|---|---|---|---|
| | | 15th %ile | 50th %ile | 85th %ile | |
| Urban | Urban 2 | 39 | 106 | 253 | 100% |
| | Urban 3 | 67 | 164 | 369 | 100% |
| | Urban 4 | 25 | 101 | 329 | 40m per 200m |
| | Rural 1 (urban only) | 83 | 195 | 477 | 50m per 500m |
| | Rural 2 (urban only) | 43 | 159 | 381 | Major arterials 100%<br>Minor arterials 100m/200m<br>Collectors 50m /200m<br>Local roads 20m/200m |
| Urban/rural | Rural 1 (combined) | 106 | 289 | 1370 | 50m per 500m |
| | Rural 2 (combined) | 65 | 306 | 1325 | Major arterials 100%<br>Minor arterials 100m/200m<br>Collectors 50m/200m<br>Local roads 20m/200m |
| Rural | Rural 1 (rural only) | 220 | 1015 | 2480 | 50m per 500m |
| | Rural 2 (rural only) | 258 | 957 | 1808 | As above |
| | Rural 3 | 76 | 616 | 2121 | 50/500 for < 500vpd<br>20/200 for > 500vpd |
| | Rural 4 | 131 | 500 | 1933 | 50m per 500m |
| State highway | SH4 | 56 | 240 | 819 | Min 50m per 500m |
| | SH6 | 128 | 313 | 659 | Min 50m per 500m |
| | SH7 | 105 | 293 | 773 | Min 50m per 500m |
| | SH8 | 90 | 239 | 596 | Min 50m per 500m |

Table 6.4 shows that urban networks have shorter treatment lengths than rural networks, as would be expected. Also, the treatment lengths are reasonably consistent across the state highway networks.

The rural networks show the greatest length. At present, many of these networks are likely to be rated using a 50m sample length per 500m resulting generally in one rated length per treatment length. As the data above shows, this gives a very poor correlation between the estimated and actual fault values. To improve the quality and consistency in the data the sample length should be decreased and frequency increased. This will give data that better represents the condition of the treatment length.

Urban networks are generally rated with a higher sample percentage than rural networks, with many rating 100%. This should produce data that we can be confident in. Even for lesser sampling regimes, the short treatment lengths usually give a high proportion of the treatment length being inspected.

The state highway network is similar to a rural network with the majority of treatment lengths having a single rating, although the lengths are shorter. The state highways are rated at 50m per 500m. However the majority of the treatment lengths are less than 500m long and therefore the percentage sample size is typically 13%–18%. A more frequent sampling at 20m or 40m per 200m would improve results significantly for both forward work programming and consistency of network condition indices.

## 6.2 What is the impact of assessments at different times of the year?

It is anticipated that some visually rated faults will vary depending on the time of year surveyed. Table 6.5 explores these reasons.

**Table 6.5 Theoretical seasonal impact of visual rated parameter**

| Parameter | Seasonal effect |
|---|---|
| Alligator cracking | Alligator cracking should be more easily identified during the colder winter months when the crack widths are greater due to the contracting pavement surfacing.<br>Also, warmer temperatures can cause ‘self-healing’ of the bitumen. |
| Potholes and pothole patches | Pothole and patch numbers are expected to be higher during winter months, particularly those networks subjected to repeated freeze/thaw cycles and higher rainfall, increased water ingress and higher water tables. |
| Shoving/rutting | Shoving and rutting are likely to be more prevalent in the winter as per potholes above. |
| Flushing | Flushing is expected to be more common during the warmer summer months when the bitumen is more fluid. |

To determine whether there are any seasonal variations in the rating data we analysed three consecutive RCA surveys, carried out in both summer and in winter (see table 6.6).

**Table 6.6 Variations in CI and PII values including some of the visual rating inputs between seasons**

| RCA | CI value | | PII value (w/o roughness) | | ACA | | ASH | | APT | | APH | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Winter | Summer | Winter | Summer | Winter | Summer | Winter | Summer | Winter | Summer | Winter | Summer |
| Urban 4 | 4.89 | 2.19 | 2.06 | 0.79 | 0.40 | 0.14 | 0.00 | 0.01 | 0.03 | 0.01 | 0.00 | 0.00 |
| Rural 2 | 0.63 | 4.82 | 1.27 | 3.55 | 0.21 | 0.75 | 0.08 | 0.10 | 0.01 | 0.00 | 0.00 | 0.00 |
| Rural 3 | 1.56 | 2.23 | 0.68 | 1.27 | 0.12 | 0.19 | 0.05 | 0.13 | 0.00 | 0.00 | 0.00 | 0.01 |

Table 6.6 shows the CI and PII values (excluding roughness) for the three networks and the contribution by those parameters expected to show seasonal variations for a summer and winter survey. The data shows that two of the three networks have larger CI and PII values (excluding roughness) for the summer surveys. Alligator cracking is probably the greatest cause of this.

This is not what was expected and as the level of change for each of the three networks is relatively small, the results are inconclusive. There may be seasonal impacts from carrying out the rating surveys at different times of the year but these are likely to be outweighed by some of the issues identified in the study.

Overall, this may be difficult to analyse. A higher number of faults in winter results in increased maintenance activity in this period which upsets the possibility of establishing consistent patterns. Furthermore, should rating be confined to a consistent window in the year, this would impact considerably on retaining rating staff during the off-season, plus increasing resourcing demands in the window period. It could be more beneficial to maintain an experienced workforce with a consistent workload throughout the year than employ a short-term, large number of less experienced raters to account for seasonal variations.

However, it is strongly recommended that individual RCAs undertake condition rating surveys consistently at the same time of year.

## 6.3 Effects of different speeds

To establish the effect of survey speed on the consistency of the data we reviewed RAMM rating survey against LTPP site surveys. The LTPP site surveys are much more controlled and detailed taking a greater length of time to complete and therefore should produce more accurate and consistent data.

Nine LTPP sites were rated in accordance with the PFM 6 and compared with the LTPP survey results. The faults recorded were analysed to establish the level of variation in the data and any consequences this would cause. It was difficult, however, to get comparable results due to the differences in fault definition.

### 6.3.1 Alligator cracking

The results of the alligator cracking recorded show that this fault type is generally under-reported when collected during the RAMM visual rating survey (see figure 6.1). There are also a large number of instances where nothing was recorded in the RAMM survey but identified during the LTPP survey. Some of this can be attributed to the RAMM survey being undertaken from the shoulder while the LTPP survey is undertaken when the lane is closed off. There are some instances where alligator cracking was recorded during the

RAMM survey but not the LTPP survey. This could be a result of misidentification during the RAMM condition survey.

**Figure 6.1 A comparison of the distribution of fault values in the RAMM and LTPP surveys**

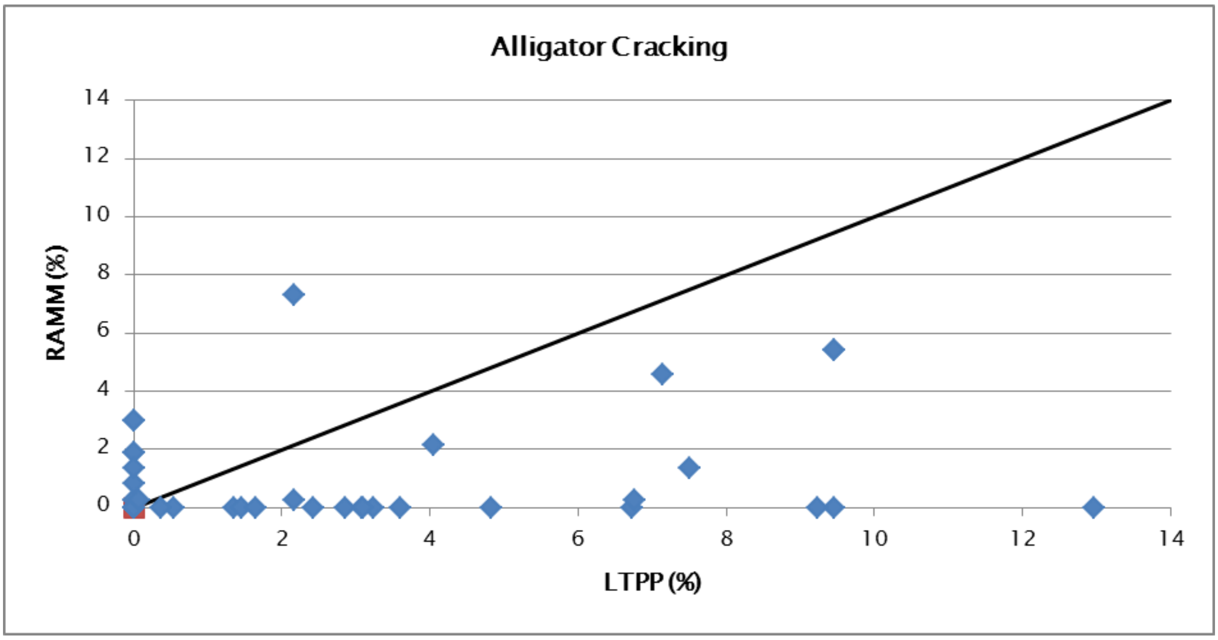


**Table 6.7 TSA proposed treatment comparison for RAMM and LTPP alligator cracking survey data**

| Site | RAMM (WPL) | LTPP (WPL) | Length | RAMM (%WPL) | LTPP (%WPL) | RAMM TSA treatment | LTPP TSA treatment |
|---|---|---|---|---|---|---|---|
| LTPP 1 | 0 | 0 | 300 | 0% | 0% | General maintenance | General maintenance |
| LTPP 2 | 0 | 48 | 300 | 0% | 4% | General maintenance | Reseal year 1 |
| LTPP 3 | 112 | 93 | 450 | 6% | 5% | Reseal year 1 | Reseal year 1 |
| LTPP 4 | 0 | 5 | 300 | 0% | 0% | General maintenance | General maintenance |
| LTPP 5 | 1 | 37.8 | 300 | 0% | 3% | General maintenance | Reseal year 1 |
| LTPP 6 | 8 | 18.67 | 300 | 1% | 2% | General maintenance | Reseal in next budget year |
| LTPP 7 | 0 | 108.8 | 300 | 0% | 9% | General maintenance | Reseal year 1 |

Alligator cracking is one of the visual rating parameters that trigger a treatment when running TSA and dTIMS. To assess the implications of this data spread we compared the proposed TSA treatment for each site. Table 6.7 shows a large inconsistency between the TSA outputs based on the survey data for alligator cracking recorded during a RAMM visual rating survey and the data from the LTPP site survey. Only 33% of the sites have consistent treatments. For the other sites the RAMM visual survey results give a lesser treatment with no requirement for a reseal in the next two years.

From this, it is evident that the faster RAMM visual rating survey is under-reporting, and on a number of sites is missing alligator cracking faults that were identified during the LTPP site survey. This is concerning as carriageway sections requiring treatment may not be identified by the RAMM visual rating survey and could cause increased maintenance costs and an inaccurate FWP.

## 6.4 What methodologies can be best used to provide quality assurance in the data?

### 6.4.1 Improved methodologies

The current approved road condition rating course is run annually by NZIHT. It is a two-day course for new raters; accredited raters attend the second day only as a refresher workshop every two years. There is an assessment but the limits of variation are such that it is very difficult to fail. It is recommended that the assessment criteria is tightened so that raters complete the course with a consistent approach to rating. The opportunity for feedback to raters may need to be improved, perhaps through limiting numbers to allow better individual assessment and feedback.

It should be noted, however, that securing consistency in survey results requires appropriate and effective QA systems for field surveys as well as improved training.

The PFM 6 requires updating with better guidance on fault definition and photographs of fault types, particularly of different types of cracking.

The PFM 6 currently requires the identification of a 5% validation area to ensure accuracy of the data collected. It is recommended the manual is amended to include QA practice guidelines covering:

- data audit checks to be undertaken prior to commencing survey
- how to obtain a 5% sample, ie selecting sections with faults
- the use of an independent auditor, ie survey team members should not check each other’s work
- an independent audit process including:
    - the rater repeating the audit in the presence of the auditor
    - a separate rating by an independent auditor and a comparison of the results
- the use of common rating sites (see below)
- data audit checks to be undertaken prior to loading in the RAMM database
- the implementation of a common rating part-sample surveyed by all raters, including the QA person, to identify any consistent under or over-reporting. An assessment would need to be made on the level of variance that would require intervention. The limits of variation would not need to apply. This could be a measure based on standard deviation from the mean for each rater. An overall standard deviation of results could also be used as a confidence measure of the survey data
- tighter limits of variation as discussed in section 4.5.2
- roads selected for QA purposes should have a high proportion with faults present to enable any issues to be identified. This could include sections with faults from previous surveys, or sites on the basis of surface age. Random selection of sites could lead to a high proportion with no faults which is not ideal
- the implementation of an independent validation service to achieve a better consistency across networks, for example, an NZTA-funded audit team, or an alternative supplier to do a sample on certain networks or road hierarchies of key importance.

A draft methodology for QA of road rating is attached in appendix A, which gives a suggested format for QA processes including field and data entry audits.

The results of the QA validation should be presented by the service provider showing compliance with the limits of variation and detailing any corrective actions required.

### 6.4.2 Trial results

The process was trialled on the urban 2 network road condition rating surveys. Figure 6.2 shows the results of two raters' reporting for alligator cracking compared with the independent auditor's rating. The shaded area represents the allowable limits of variation.

Rater 1 shows a trend of under-reporting the cracking faults compared with the QA auditor. This pattern is very difficult to detect by just selecting sections outside the limits of variation. Here the sites were selected on the basis of there being a high proportion with faults. Normally only one or two sites would be outside the limits of variation.

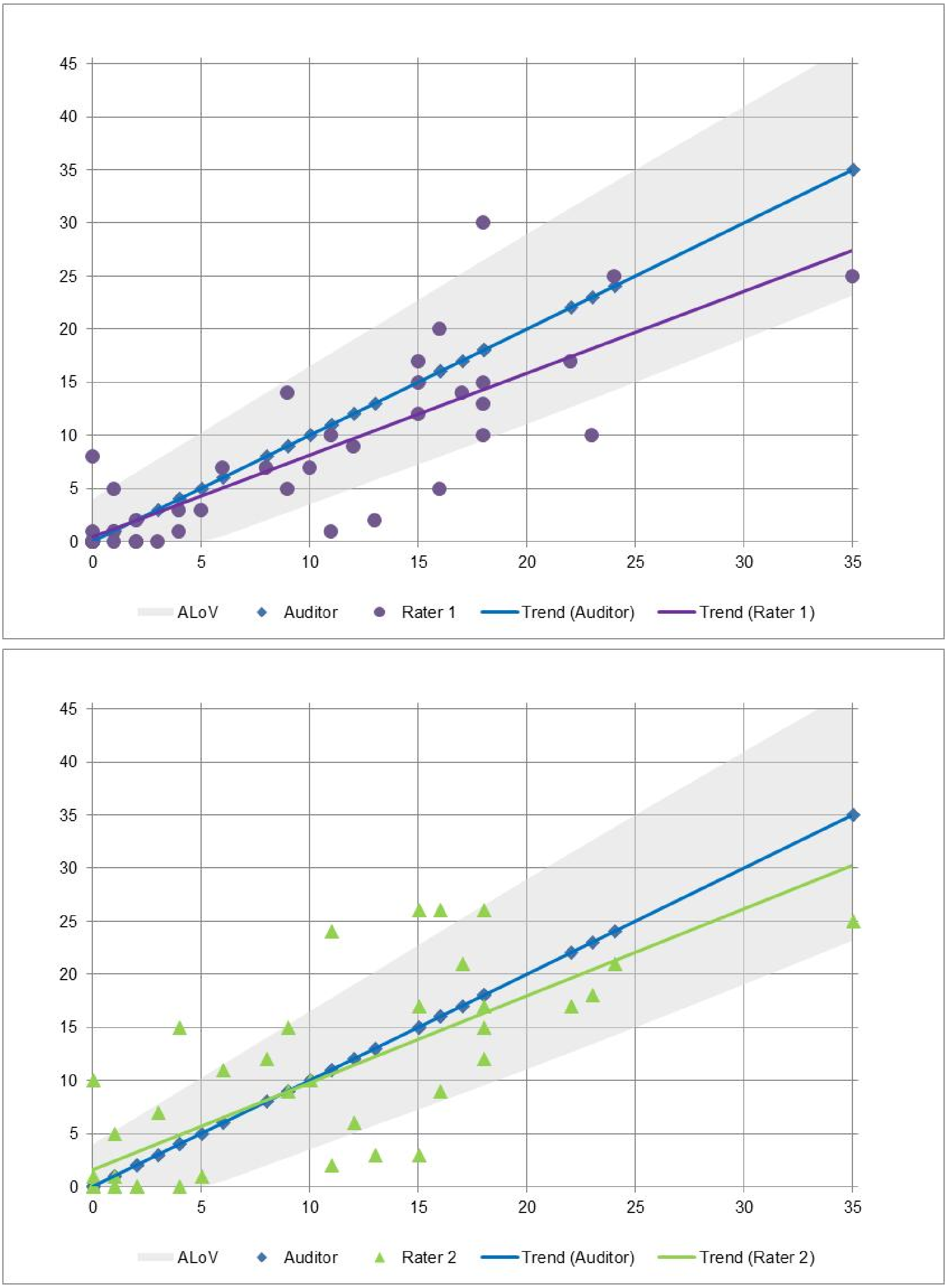
In this case, the sites outside the allowable limits of variation would be resurveyed. This provides a good opportunity to identify where the rater was not being consistent and to discuss with them their under-reporting of faults. It may be that they were in too much of a hurry or their estimation of the length of the fault was too low. This problem can be addressed and a more consistent result achieved.

Rater 2 has more scatter in their result but the overall trend is similar to that of the auditor. The auditor will use these results to discuss with the raters how they can achieve more consistency.

This analysis provides a clear picture of rater performance and areas for improvement. It should also be noted that this activity takes place continuously as surveys are completed. Therefore a rater's on-going performance can be monitored. Also, one would expect the consistency across the rating team to improve over time as the feedback builds up.

This process will result in greater consistency of rating data and less discrepancy between different raters and rating teams.

**Figure 6.2 Comparison of trial common rating QA sections**



## 6.5 Measures to prioritise rigour in the data collection process

As identified in section 6.1.1, reducing the sample size and sample frequency results in variations in the index values and treatment selection options due to under-reporting of the faults. This means that any stratification can produce similar variations across networks. Therefore a minimum standard of survey frequency needs to be implemented to minimise this effect.

A maximum sample length should be 200m. A minimum 20m inspection length would provide a 10% sample, while on higher volume roads a minimum of 40m or 20% could be used. RCAs can still choose to go to 100% sampling to provide greater confidence and accuracy. However, by adopting the proposed regime, the impact of under-reporting faults would be minimised.

This stratification could be undertaken by either traffic volume (eg greater or less than 500 vpd) or by hierarchy (say local roads at 10%; arterials, collectors, etc at 20%).

The use of 200m sections would not require any split between urban and rural stratification.

A second question is the frequency of surveys. One of the difficult factors at the moment is the minimum requirement of surveys every two years, but the NZTA funding cycle is every three years. There are currently the following options:

- annual surveys, but this would be inefficient
- main roads (set by traffic volume or hierarchy) surveyed annually and secondary roads surveyed every two years:
    - the entire secondary network could be done every two years
    - half the secondary network could be done annually
- a minimum of biennial surveys of the entire network with RCAs choosing what best suits their needs.

There has been some debate regarding the splitting of surveys and the impact on condition index reporting. As long as the latest survey is taken, this should not be an issue as the entire network condition will be reported each year. A review of the data for the rural 4 network, where 50% of the secondary roads are rated each year, showed negligible impact fault trend graphs.

We recommend that where secondary roads are to be surveyed every two years, the entire network is surveyed every second year.

Consideration should also be given to the timing of the survey. It should align with the RCA's planning process which would mean carrying it out prior to when the treatment selection and FWP takes place.

## 6.6 Options for a confidence level system on existing data

We recommend using the overlap rating system to provide a measure of variability and therefore confidence in the survey data required. The implementation of a common rating part-sample of the network surveyed by all raters and the QA auditor, is to identify any consistent under or over-reporting. A measure based on standard deviation of overall results could be used as a confidence measure of the survey data. These sections could also be rated by an independent party to provide a confidence measure.

## 6.7 Procurement

Pradhan (2009) indicated one third of TLAs use a single year procurement term, although on anecdotal evidence this seems high. Pradhan makes the point that longer-term contracts provide consistency as the same team is used each year. We concur with this view. This method also reduces tendering costs.

We recommend the following for stand-alone condition assessment contracts:

- multi-year, preferably 3+1+1 term contracts, or 4+2 if roads are surveyed on alternate years
- weighted attribute rather than lowest price selection
- specified QA requirements, preferably based on best practice guidelines as part of an updated rating manual
- a single combined contract for smaller networks. Hurunui, Waimakariri and Kaikoura districts have successfully operated a similar system for a number of years. It will produce efficiencies for the RCAs but will probably have a limited impact on results.

In our experience the contract documentation is generally standard, particularly concerning rating requirements, deliverables, and limits of variation and calibration procedures. However, the requirements for QA are varied. The QA requirements can be prescribed, left to the rating team or not mentioned at all. It is recommended that guidance is provided for QA procedures. A proposal can be found in appendix A.

There may be benefits and cost savings in combining the road condition rating surveys with other road asset management activities such as RAMM management, TSA delivery, dTIMS, FWP, etc. However, this may be lost at the tender box as a result of a reduced market capability to deliver all these activities.

# 7 Conclusions and recommendations

## 7.1 Conclusions

### 7.1.1 Fault measurement

Table 7.1 summarises the impact of the visually rated parameters on their current use. For the performance measures SCI and PII, the quality of the data for alligator cracking, potholes, shoving, flushing and scabbing needs to be good. For treatment selection and modelling, alligator cracking, potholes, pothole patches and flushing were found to have the greatest influence on the outcomes.

**Table 7.1 Impact of visual rating parameters on performance measures and proposed improvements**

| Parameter | Frequency | Typical value | Spread of data | Impact on SCI | Impact on PII | Impact on TSA | Impact on dTIMS | Proposed action |
|---|---|---|---|---|---|---|---|---|
| Alligator cracking | Medium | Small | Poor | High | Medium | Medium | High | Tighten limits of variation. Improve fault identification |
| Potholes | Low | Small | Good | Medium | Negligible | Medium | Medium | OK currently |
| Pothole patches | Low | Small | Good | Low | Negligible | Medium | Medium | OK currently |
| Shoving | Low | Small | Good | - | Medium/high | Low | Medium | Tighten limits of variation.<br>Collect by HSDC on higher risk roads. |
| Rutting | Low | Small | Good | - | Low | - | High | Collect by HSDC on higher risk roads.<br>Change manual rating to method consistent with HSDC. |
| Flushing | Low | Small | Good | Medium | - | Medium | High | Change to Cat A, Collect by HSDC? |
| Scabbing | High | Small | Poor | Medium | - | Low | Medium | Change to Cat A with tighter limits of variation. |

### 7.1.2 Alligator cracking

For alligator cracking, the distribution of raters' values during the annual rating course was found to be poor. The quality of data recorded for this parameter is important as it creates confidence in the results of the performance measures, the TSA and modelling. A visual rating survey is currently the best method for identifying alligator cracking; however, measures need to be established to improve accuracy and confidence.

The current acceptable limits of variation need to be tightened to encourage the reduction of under or over-reporting of this fault type to tolerances that meet the requirements of the outputs for which the data is used. As the current tolerance was found to be exceeded regularly during the rating course, improvements are needed in the identification and capture of this fault type.

This could be achieved through updating the PFM 6 to include better examples, giving greater emphasis to this fault at the annual rating course as well as undertaking further rating inspections following the comments received during the initial inspections, and ensuring a robust QA procedure is identified and implemented.

### 7.1.3 Shoving

Shoving can be collected by HSDC. This automated method should provide much more consistent data although the visual condition rating data is reasonably consistent. An automated data capture survey will be more expensive than the visual rating surveys but the quality and speed at which the data can be collected will be improved. This could be combined with the roughness survey to improve efficiency and reduce costs.

HSDC would also remove a number of fault types recorded during the visual rating surveys, enabling better identification of other parameters, particularly alligator cracking.

### 7.1.4 Rutting

Rutting can also be collected by HSDC. This automated method should provide much more consistent data. An automated data capture survey will be more expensive than the visual rating surveys but the quality and speed at which the data can be collected will be improved.

It is recommended that the rutting data be collected by assessing the length of wheelpath to the nearest 5mm, which is a similar method to the one given in the rating manual. This would create a methodology consistent with HSDC and would be less of an issue when comparing network condition and indices where different data collection methods are used.

The introduction of HSDC for rutting, combined with the visual condition rating was trialled by Hastings District Council. This method gave a more strategic approach to targeting HSDC at key routes and the cost was offset by a reduction in the frequency of roughness surveys on lower volume roads. This proved successful and provided a much improved strategic analysis process for the council's FWP and asset management plan.

### 7.1.5 Flushing

Flushing is currently a category B fault type. It therefore has a greater tolerance than those listed as category A. To have confidence in the data, particularly when used in the SCI value calculation and treatment selection, it should be changed to category A.

Like shoving and rutting, it is possible to capture flushing using automated HSDC which should produce more accurate and consistent data.

### 7.1.6 Scabbing

Scabbing like flushing is currently a category B fault type. Because of the impact it has on the SCI value calculation scabbing should be changed to category A to improve data quality.

Scabbing was also found to have a wide distribution of rater values at the annual rating course. Like alligator cracking this could be addressed through improvements to the PFM 6, the annual rating course and implementing a robust QA procedure.

## 7.2 Recommendations

### 7.2.1 Rating manual changes

The manual requires updating as follows:

- provide improved guidance on fault definition
- include photographs of fault types, particularly of different types of cracking
- establish a new category for the fault types with a high influence on the outputs for which they are used (alligator cracking, shoving, rutting and flushing).
- tighten the limits of variation for category A to:

$$L = \pm 1.2 \times \sqrt{Va} \qquad \text{where } Va > 12 \qquad \text{(Equation 7.1)}$$

$$L = \pm (\tfrac{1}{4} Va + 1) \qquad \text{where } Va <= 12 \qquad \text{(Equation 7.2)}$$

- change to category A the defect types that are currently category B and have a significant influence on the outputs for which they are used (rutting, flushing and scabbing).

### 7.2.2 Rater training

The assessment criteria should be tightened so that raters complete the course with a consistent approach to rating. The opportunity for feedback to raters may need to be improved, perhaps through limiting numbers to allow improved individual assessment and feedback.

It should be noted, however, that the consistency of survey results should be achieved through the application of appropriate and effective QA systems. It is not the role of the training course to provide the industry with raters ready to perform rating to a consistent and high standard with no further training and minimal QA monitoring. The course is to equip raters with the skills and training they need to be able undertake the condition rating surveys. It is the role of the organisations employing the raters to give the further training, consistent feedback and monitoring through the QA process, and to provide accurate and robust data.

### 7.2.3 QA procedures

The PFM 6 currently requires the identification of a 5% validation area to ensure accuracy of the data collected. It is recommended the manual is amended to include QA practice guidelines covering:

- data audit checks to be undertaken prior to commencing survey

- how to obtain a 5% sample, ie selecting sections with faults
- the use of an independent auditor, ie survey team members should not check each other's work
- an independent audit process including:
    - the rater repeating the audit in the presence of the auditor
    - a separate rating by an independent auditor and a comparison of the results
- use of common rating sites (see below)
- data audit checks undertaken prior to loading in the RAMM database
- the implementation of a common rating sample surveyed by all raters, including the QA person, to identify any consistent under or over-reporting. An assessment would need to be made on the level of variance that would require intervention. The limits of variation would not need to apply. This could be a measure based on standard deviation from the mean for each rater.
- tighter limits of variation as discussed in section 4.5.2
- roads selected for QA purposes should have a high proportion with faults present to enable any issues to be identified. This could include sections with faults from previous surveys, or sites on the basis of surface age. Random selection of sites could lead to a high proportion with no faults which is not ideal
- the implementation of an independent validation service to achieve a better consistency across networks, for example, an NZTA-funded audit team, or an alternative supplier to do a sample on certain networks or road hierarchies of key importance.

The results of the QA validation should be presented by the service provider showing compliance with the limits of variation and detailing any corrective actions required.

### 7.2.4 Stratification and sampling

A maximum sample length of 200m is recommended. A minimum 20m inspection length would provide a 10% sample while on higher volume roads, a minimum of 40m or 20% could be used. RCAs can still choose to go to 100% sampling to provide greater confidence and accuracy. However the impact of under-reporting faults will be minimised.

This stratification could be undertaken by either traffic volume (eg greater or less than 500 vpd) or by hierarchy (say local roads at 10%; arterials, collectors, etc at 20%).

The use of 200m sections would not require any split between urban and rural road sections.

One of the difficult factors at the moment is the minimum requirement of surveys every two years, but the NZTA funding cycle is every three years. There are currently the following options:

- annual surveys, but this would be inefficient
- main roads (set by traffic volume or hierarchy) surveyed annually and secondary roads surveyed every two years:
    - the entire secondary network could be done every two years

  - half the secondary network could be done annually
- a minimum of biennial surveys of the entire network with RCAs choosing what best suits their needs.

It is recommended that RCAs undertake condition rating surveys consistently at the same time of year.

The use of HSDC on higher-level roads is recommended and discussed in more detail in section 7.1.4 on on rutting.

### 7.2.5 Confidence level system on existing data

The implementation of a common rating sample, (appendix A) of the network which is surveyed by all raters including the QA person, will identify any consistent under or over-reporting. A measure based on the standard deviation of overall results could be used as a confidence measure of the survey data. These sections could also be rated by an independent party to provide a confidence measure also.

### 7.2.6 Procurement

Longer-term contracts provide consistency as the same team is used each year. This method also reduces tendering costs.

We recommend the following for stand-alone condition assessment contracts:

- multi-year, preferably 3+1+1 term contracts, or 4+2 if roads are surveyed on alternate years
- weighted attribute rather than lowest price conforming
- specified QA requirements, preferably based on best practice guidelines as part of an updated rating manual
- a single combined contract for smaller networks. Hurunui, Waimakariri and Kaikoura districts have successfully operated a similar system for a number of years. It will produce efficiencies for the RCAs but will probably have a limited impact on results.

Documentation is consistent in our experience although QA practice can vary. However the documentation is generally standard, particularly concerning rating requirements, deliverables, and limits of variation and calibration procedures. An improved guide to QA procedures in the rating manual will assist with this.

There may be benefits and cost savings gained in combining the road condition rating surveys with other road asset management activities such as RAMM management, TSA delivery, dTIMS, FWP, etc. However, this may be lost at the tender box as a result of a reduced market capability to deliver all these activities.

# 8 References

Beca Consultants (1997) *An investigation into rating survey sample length correlation.* A report prepared for Transfund New Zealand.

Bennett, CR (2001) Evaluating the quality of road survey data. *Transfund NZ research report 2001*.

CJN Technologies (2005) *Working with treatment lengths*. Version 1.0.

Committee of State Road Authorities (1992) *TMH9: Pavement management systems: standard visual assessment manual for flexible pavements*. Pretoria: The Department of Transport.

Deng, Z and TFP Henning (2013) *Testing of new NZTA condition performance measures*. Wellington. NZTA.

Fwa, TF, SB Liu and KJ Teng (2003) Airport pavement condition rating and maintenance needs assessment using fuzzy logic. Airfield pavements: challenges and new technologies. In *Proceedings of Airfield Pavements Specialty Conference* 2003. ASCE.

Haas, R (2001) Reinventing the (pavement management) wheel. *5th Annual Conference on Managing Pavements*, Seattle, Washington, USA, 2001.

Hastings District Council (HDC) (2013) *Transportation asset management plan*. Hastings: Hastings District Council.

Henning, TFP, SB Costello, CC Parkman, CR Bennett and RCM Dunn (2006) A complete review of the crack initiation models for New Zealand. *22nd ARRB Conference Research into Practice*, Canberra, Australia, 29 October – 2 November 2006.

Henning, TFP (2009) *The development of pavement deterioration models on the state highway network of New Zealand*. Accessed 20 January 2013. http://researchspace.auckland.ac.nz/handle/2292/4236

HTC (2000) *Data issue report*. Auckland: HTC Infrastructure Management Ltd.

Infrastructure Decision Support (IDS) Ltd (2008). *Reviewing the SII and PCI indices*. Report prepared for the NZ Transport Agency. 41pp.

New Zealand Transport Agency (NZTA) (2011) *Planning and investment knowledge base*. Wellington: NZTA.

New Zealand Transport Agency (NZTA) (2008) *State highway national pavement condition report*. Wellington: NZTA.

New Zealand Transport Agency (NZTA) (2009) *State highway national pavement condition report*. Wellington: NZTA.

Perera, S (2010) *The applicability of newly developed structural indices for road conditions reporting, monitoring and pavement deterioration modelling.* Auckland: University of Auckland, Department of Civil and Environmental Engineering

Piane, R (2010) *Case studies of dTIMS uses in other countries*. Deighton presentations during IDS/Deighton Workshops. Auckland.

Pradhan, N (2009) *Review of pavement condition rating practices. Report no. PM07/1264T*. Wellington: NZTA.

Rainsford, S (2004) *Performance condition index (PCI) national investigation*. RIMS Group.

Transfund New Zealand (1997) *RAMM road condition rating and roughness manual* (PFM6). Wellington: Transfund NZ. 66pp.

Transit New Zealand (2000) *Calibrations section survey specifications*, Wellington: Transit New Zealand.

# Appendix A: Proposed model quality assurance requirement

## A1 Validation process and on-going quality assurance

### A1.1 Field work Instruction

Prior to commencement of the survey each year, the Contractor shall forward the field work instruction. The field work instruction is given to the rating field team and details the specific requirements for the survey. It shall include the following as a minimum:

- scope of the survey
- Client requirements including any special Client requirements that vary the standard rating manual requirements, roads with restricted access etc
- calibration strip locations
- approved traffic management plan
- team structure and assigned work areas
- any contract specific issues such as dealing with the public
- list of hazards that may be encountered and mitigating measures.

### A1.2 Rating validation

To ensure the accuracy and consistency of the data collected, the contract requires a rating validation process involving the Client and Contractor and/or their representatives using 100 rating sections for the road condition rating. The sections to be validated shall be chosen by the Client. The validation process shall take place prior to rating commencing and shall involve the following:

- the Contractor's quality control auditor
- the Contractor's rating surveyors
- the Client's relevant engineering staff
- any independent party the Client may wish to engage.

As a group, these individuals will travel to the Client's nominated sections where the group members will individually rate the same sections. The results will be compared at the end of each exercise for the purpose of reaching consistency in ratings. This may require re-rating of sections to achieve consistency.

All raters and quality assurance (QA) audit staff who are to rate the main survey are required to rate the validation sites during the validation week, unless otherwise agreed with the Client.

## A1.3 On-going QA for rating (Client)

Once the rating validation (see above) is complete, the Client may from time to time arrange random rating audits with the Contractor's quality control auditor. If there are any survey rating forms where more than two faults are found outside the limits of variation, or if there are consistent gross or repetitive errors, the Contractor's rating team which produced the error(s) will be required to undergo training in the faults in question, and then re-survey all the sections where data fields were found to be inaccurate. There shall be no addition to the time taken to complete this Contract consequential to these requirements.

No payment shall be made to the Contractor until the Client is satisfied with the accuracy and completeness of the rating survey(s) requested on the relevant purchase order(s).

Failure to complete the condition rating to specified standards within four weeks from being notified by the Client of inaccuracies, or incompleteness, will be grounds to terminate this part of the Contract without payment to the Contractor for the faulty work.

## A1.4 On-going QA for rating (Contractor)

During the course of the survey, the Contractor shall review a minimum 5% sample of the total rating sections completed. The selection will be agreed with the Client such that it represents sites with a spread of condition to enable assessment of rater performance in assessing fault quantities.

Two percent of these sites shall be resurveyed independently by the Contractor's nominated quality control auditor. The result of this survey and the original survey by the rater shall be compared against the limits of variation.

Two percent of sites shall be resurveyed by the rater (ie repeating the initial rating inspection) with the Contractor's nominated quality control auditor in attendance. This is so the auditor can observe the rater in undertaking the condition survey and address any issues observed. The original survey and repeated survey shall then be compared against the limits of variation.

One percent of sites shall be common sites which are independently rated by the Contractor's nominated quality control auditor and each member of the Contractors rating team. This shall be completed through the rating survey to ensure the rating team's condition assessments remain constant relative to each other.

For data entry QA, 5% of forms shall be randomly selected and the fields checked that the data has been loaded into RAMM correctly. (This may not be required for data entered directly into data loggers.)

All data loaded shall also have audit checks run to identify any out of range entries or missing fields and these updated and corrected.

Following submission of data to the Client for each of the annual rating surveys, the Contractor is required to provide a quality report to the Client detailing:

- number of rating sections checked by the quality control auditor
- number of rating sections found to have two or more items outside the limits of variation, and the remedial measures taken
- number of rating sections found to have gross or repetitive errors and the remedial measures taken
- for data entry, the number of rows checked and the number of errors found and corrected.

**Table A.1 Road condition rating table and data audit recommendations**

| Field | Description | Data entry audit range |
|---|---|---|
| **Road name** | | |
| Road ID | RAMM road ID | Not null |
| Carriageway start | Start displacement in metres from the road origin | 0 – 10,000 |
| Start | Start displacement (m) of rating section from the road origin | 0 – 10,000,<br>>= Cway start |
| End | End displacement (m) of rating section from the road origin | 0 – 10,000,<br>>= Cway start |
| Date | Date of rating survey | Within survey date range |
| Inspection start | Start displacement (m) of inspection length from the road origin | >= Start<br><= End |
| Inspection end | End displacement (m) of inspection length from the road origin | >= Start<br><= End |
| Survey number | The unique number the identifies the survey | Matches survey header |
| Latest | Latest rating section or not | |
| No. of lanes | Number of TRAFFIC LANES (inspection length) | < 5 |
| Broken channel LHS | Length INEFFECTIVE due to the BROKEN CHANNEL (rating length) | < 1000 |
| High lip LHS | Length INEFFECTIVE due to HIGH CHANNEL LIP (rating length) | < 1000 |
| Broken surface LHS | Length INEFFECTIVE due to BROKEN C/W SURFACE at channel (rating | < 1000 |
| Blocked channel LHS | Length INEFFECTIVE due to BLOCKED CHANNEL (rating length) | < 1000 |
| Uphill grade LHS | Length INEFFECTIVE due to UPHILL GRADE (rating length) | < 1000 |
| Blocked earth channel | Length of BLOCKED EARTH CHANNEL (rating length) | < 1000 |
| Inadequate earth channel | Length of INADEQUATE EARTH CHANNEL (rating length) | < 1000 |
| Ineffective shoulder LHS | Length of SHOULDER that CANNOT SHED WATER (rating length) | < 1000 |
| Broken channel RHS | Length INEFFECTIVE due to the BROKEN CHANNEL (rating length) | < 1000 |
| High lip RHS | Length INEFFECTIVE due to HIGH CHANNEL LIP (rating length) | < 1000 |
| Broken surface RHS | Length INEFFECTIVE due to BROKEN C/W SURFACE at channel (rating | < 1000 |
| Blocked channel RHS | Length INEFFECTIVE due to BLOCKED CHANNEL (rating length) | < 1000 |
| Uphill grade RHS | Length INEFFECTIVE due to UPHILL GRADE (rating length) | < 1000 |
| Blocked earth channel | Length of BLOCKED EARTH CHANNEL (rating length) | < 1000 |
| Inadequate earth channel | Length of INADEQUATE EARTH CHANNEL (rating length) | < 1000 |
| Ineffective shoulder RHS | Length of SHOULDER that CANNOT SHED WATER (rating length) | < 1000 |

| **Field** | **Description** | **Data entry audit range** |
|---|---|---|
| **Road name** | | |
| SWC severity | Surface water channel severity indicator for the rating section | < 1000 |
| Rutting | Length of WHEELPATH RUTTING > 30mm (inspection length) | < 1000 |
| Rut mean depth | Mean depth of rutting within the inspection length | < 20 |
| Rut mean depth stddev | Standard deviation of the rutting mean depth | < 20 |
| Shoving | Length of SHOVING (shallow shear) (inspection length) | < 1000 |
| Scabbing | Area of SCABBING (> 10% stone loss) (inspection length) | < 1000 |
| Flushing | Length of WHEELPATH FLUSHING (inspection length) | < 1000 |
| Alligator cracks | Length of WHEELPATH ALLIGATOR CRACKING (inspection length) | < 1000 |
| L and T cracks | Length of LONGITUDINAL & TRANSVERSE CRACKING (inspection length) | < 1000 |
| Joints | Length of JOINT CRACKS (inspection length) | < 1000 |
| Potholes | Number of POT HOLES (inspection length) | < 1000 |
| Pothole patches | Number of POT HOLE PATCHES (inspection length) | < 1000 |
| Edge break | Length of EDGE BREAK (> 100mm) if no surfaced SWCs (inspection length) | < 1000 |
| Edge break patches | Length of EDGE BREAK PATCHES if no surfaced SWCs (inspection length) | < 1000 |
| Service covers | Number of service covers more than 10mm above/below the seal | < 1000 |
| Service trenches | Number of service trenches more than 10mm above/below the seal | < 1000 |
| Maintenance patches | Area of maintenance patches (include maintenance patches where > 1m2 in area) | < 1000 |
| Rater | Contract name plus the initials or name of the rater | |
| Notes | General comments | |
| Date added | The date this row was added | |
| Added by | The login name of the person who added this row | |
| Date changed | The date this row was last changed | |
| Changed by | The login name of the person who last changed this row | |
| Rating ID | Rating section ID number | |