Evaluating the Quality of Road Survey Data

Transfund New Zealand Research Report No. 200

Evaluating the Quality of Road Survey Data

C.R. Bennett
HTC Infrastructure Management Ltd,
PO Box 177, Kumeu, Auckland 1250

Transfund New Zealand Research Report No. 200

ISBN 0-478-25058-4
ISSN 1174-0574

PO Box 2331,Lambton Quay, Wellington, New Zealand
Telephone 64-4 473-0220; Facsimile 64-4 499-0733

Bennett, C.R. 2001. Evaluating the quality of road survey data.
Transfund New Zealand Research Report No. 200. 96 pp.

Keywords: condition rating, data quality, databases, exception reporting, RAMM, outlier, pavements, roads, roughness, statistics, surveys, time series, variability

An Important Note for the Reader

The research detailed in this report was commissioned by Transfund New Zealand.

Transfund New Zealand is a Crown entity established under the Transit New Zealand Act 1989. Its principal objective is to allocate resources to achieve a safe and efficient roading system. Each year, Transfund New Zealand invests a portion of its funds on research that contributes to this objective.

While this report is believed to be correct at the time of publication, Transfund New Zealand, and its employees and agents involved in its preparation and publication, cannot accept any liability for its contents or for any consequences arising from its use. People using the contents of the document, whether direct or indirect, should apply, and rely upon, their own skill and judgement. They should not rely on its contents in isolation from other sources of advice and information. If necessary, they should seek appropriate legal or other expert advice in relation to their own circumstances.

The material contained in this report is the output of research and should not be construed in any way as policy adopted by Transfund New Zealand but may form the basis of future policy.

Acknowledgments

The author expresses his appreciation to the following individuals whose support made the project possible:

- Dr. Ron Allan for his reviews and comments on the work;
- Mr Zuwei Deng of HTC Infrastructure Management Ltd (HTC) for his work in preparing the project databases;
- The road controlling authorities who provided data for the project: Mr Indra Gyawali of North Shore City; Mr Alan Roland of Auckland City Council; Mr Des Parkinson of New Plymouth District Council; Mr Ian Marshall of Southland District Council; Mr Theuns Henning and Mr Dave Robertson of Transit New Zealand.

CONTENTS

Appendices:

Executive Summary

Data can be considered to be of an acceptable level of quality if they confirm to a defined specification, and that specification correctly reflects the intended usage. The objectives of this project were to establish a methodology for evaluating the quality of road survey data, and to test this on data from different road networks in New Zealand.

Road Data Collection in New Zealand

New Zealand road controlling authorities (RCAs) have annual data collection programmes for which data are collected visually and with automated equipment. Since this study of data quality was looking at attributes, which were likely to change over time, the data considered consisted of:

- Alligator Cracking
- Edge Break
- Flushing
- Potholes
- Rut Depth
- Scabbing
- Shoving
- Longitudinal and Transverse Cracking

The other distresses (joint cracks, broken or blocked channels, ineffective shoulder, surface water channel problems, and inadequate drainage) were not included in the analysis since their deterioration is not suitable for time series modelling.

High speed data measurements are comprised of roughness, rut depth and texture, although most RCAs only have roughness data. The State Highway Network has all three dating from 1994.

The annual survey data represent a time series, where the spatial location and order of the sampling points play the role of time. The goal of the project was to develop a statistical approach, and a simple method for applying it, to evaluate these data for consistency.

Errors in Data Collection

The three sources for errors in data collection are:

- random errors,
- systematic errors, and/or,
- operator errors.

If the current data do not compare well with data from the previous surveys this is due to one (or more) of four reasons:

- the current survey is wrong,
- the previous survey(s) were wrong,
- the data were collected or referenced to the wrong location, and/or,
- the pavement has been modified.

When assessing data it needs to be appreciated that each measurement has a set level of precision. Thus, when measuring the same section of road twice, there will be variation in the measurements, even if these are made with the same instrument at the same time.

An analysis of data for roughness meters suggested a standard deviation of 2% for laser profilometers and 3% for response-type meters. The implication of this is that, in order to assess pavement trends, and even evaluate data quality, one needs to have a high level of precision in the measurements. If this is not achieved the apparent pavement deterioration will be unreliable.

Assembly of Databases

In order to undertake the time series analyses of data quality, databases were assembled from six different RCAs: two State Highway; two urban local; and two rural local authorities. The source data were obtained from the RAMM databases (Road Assessment & Maintenance Management System) and these were then converted into time series wherein, for each 100-m section of road, a series of measurements had been made over a number of years.

The assembly of the databases proved to be a monumental task. This was due to a number of factors, particularly the architecture of RAMM and its failure to store data in a manner conducive to creating databases for time series analyses. There were also problems within the databases, such as poor location referencing; changes to the locations of inspection lengths; and inconsistent lane identification.

Time Trends in Pavement Condition

A particular problem was the variability of the data. When many sections were reviewed as a time series, i.e. looking at the annual changes in pavement condition for the same 100-m sections of road, often significant changes in condition were observed which could not be explained by pavement deterioration or any other factor. In general, the State Highways had much better consistency on an annual basis than local authority roads, perhaps reflecting the higher quality assurance standards of State Highway surveys or the use of the same contractor for extended periods of time.

Maintenance presented a particular problem since it leads to major changes in the condition of a pavement. Unfortunately, re-surfacings and overlays/shape corrections are not recorded reliably in RAMM, i.e. changes in pavement condition indicate that a treatment was performed but it is not in the database. The situation is more complicated with minor maintenance activities such as patching, which is only recorded through maintenance costs. In the analysis it was not possible to use the maintenance costs for identifying treatments, so only re-surfacings and overlays/shape corrections were identified.

Establishing time trends from the RAMM visual data that are suitable for statistical analysis proved to be impossible because of the impacts of maintenance, compounded by the inability to identify when minor maintenance was performed.

The failure for major maintenance (re-surfacings, shape corrections) to be reliably recorded in RAMM, or for minor activities to be adequately recorded for the purposes of time series analyses, means that "noise" will be introduced to the data which are related to unique, external factors. These lead to discontinuities in the deterioration trends which make any form of trend analysis difficult at best, and most likely meaningless in many situations since the other problems such as lack of precision and small sample sizes, are being compounded.

Analysis of Roughness Data

The roughness data were analysed using a Box-Jenkins approach to investigate whether or not assessing the validity of new data against previous time series data would be practical. The objective was to compare the current survey's value against the previous values and establish whether or not it was consistent. This would enable RCAs to quickly assess the validity of data provided by a contractor.

After considering a variety of options, a commercial application called Autobox was used for the analysis. This had a batch mode which would be suitable for large data sets. The "exception analysis" option was used to identify two types of situations which give rise to "unusual values":

- **Pulses** are unusual events in the data. For example, the measurements in one year may be incorrect due to improper calibration of the vehicle.
- **Level shifts** arise when there is a shift in the magnitude of the data, for example the roughness after shape correction will be much lower than before.

A sample data set was established from the road databases representing the full range of conditions found in the databases. The variability in the data and the small sample sizes presented a challenge to any applications analytical capabilities.

Autobox was only able to fit time series trends to 15% of the data, but this was not unexpected given the nature of the data: most did not show any viable trend, which was not unusual in some databases. In a number of instances the software successfully identified pulses and level shifts, although it also missed in a number of instances. Overall, the software was considered to have performed well considering the nature of the data. It has the potential for acting as a general data quality tool for identifying anomalies in contractor's measurements but, until the general quality of the data is improved, Autobox, or any similar tool, will struggle.

Suitability of Data for Analysis

In general, the lack of precision with much of the data in the RAMM databases, both visual RAMM rating and roughness, precludes a time series analysis for the purposes of statistically evaluating the quality of new survey data. It is possible that other analytical processes, such as using longer sample lengths or including the spatial dimension, may be better than the Box-Jenkins approach tested here.

Importantly, the need is to improve on the reliability of recording of maintenance information in RAMM. Until this is done any statistical analysis will be confused by changes in the condition which accompany road maintenance. Changes to RAMM should also be considered to make it easier to extract data for time series analyses.

If the objective is to check the quality of the data collection, certain sections can be identified as "controls", for which special effort can be made to keep accurate records. The data-collecting contractor would not know the locations of these sections, and so the resulting data would be "typical" of the overall survey quality.

Recommendations

The principal recommendations from this research are as follows:

- The way in which maintenance activities are recorded in RAMM should be reviewed with the objective of establishing a procedure which will facilitate time series analysis of the data. Other changes to the architecture of RAMM should be considered to assist with exporting the data for quality assurance evaluations.

- The precision of the various road survey measurements that are in use needs to be established. In discussing the application of data with practitioners, a precision is expected of the data, which may not in fact be appropriate. Knowing the precision of measurements under typical survey conditions would also facilitate the quality assurance process insofar as it would indicate the expected change from year-to-year related to measurement variability.

- While the time series analyses with Autobox showed that it is possible to identify unusual readings, the small sample sizes and noise in the data helped to reduce the reliability of the analysis. For practical use, too many data items would be identified as questionable using this method. It would be better to use a different approach, one which took into account the spatial element of the data, than a Box-Jenkins time series approach. Other outlier techniques are available and these should be explored. The databases developed in this project would provide the ideal springboard for developing this alternative statistical approach. The work should be developed into a stand-alone software application, which could be used by all RCAs for assessing their data quality.

Abstract

This report describes the results of a project to investigate the quality of road survey data collected from road networks in New Zealand. The objective was to establish a statistical technique which could be used to identify whether data were inconsistent with data from previous years' surveys. Seven databases were assembled with data in the format suitable for time series analyses from different road controlling authorities. This proved to be a complicated process due to problems with how the data were stored and referenced. The data were evaluated on 100-m road sections and major variations in condition between years were found which could not be ascribed to pavement deterioration. In many instances these were due to maintenance applied to the pavement that was not always recorded in the databases. The roughness data were considered to be the most suitable for time series analysis, and they were analysed using a Box-Jenkins approach with the software application Autobox. While Autobox proved suitable for identifying some of the deficiencies, it was considered that an alternative approach incorporating the spatial component and more suitable for small sample sizes would be likely to give better results.

1. Introduction

The Transfund-sponsored project *"Evaluating the Quality of Road Survey Data"* was undertaken in 2000-2001. It had the following objectives:

- To obtain a set of data on New Zealand RAMM[1] condition rating and roughness over 5 years (or more) from rural, urban and State Highway networks.
- To analyse these data and develop techniques using time series models to establish a framework for establishing whether survey data are consistent with the results of previous years.
- To establish simple yet practical guidelines for applying this method.

Since the goal of the project was to assess the quality of road data, it was first necessary to define exactly what data quality meant. Although there is broad agreement to the concept of data quality, the literature does not contain a standard definition of what this means.

Abate et al. (1998) define data to be of the required quality:

> *...if it satisfies the requirements stated in a particular specification and the specification reflects the implied needs of the user.*

Therefore, an acceptable level of quality has been achieved if the data conform to a defined specification, and the specification correctly reflects the intended use. This concept of defining data quality by conformance and utility is particularly pertinent to road data and was adopted for this project.

This report opens with a discussion of how road data are measured in New Zealand (Chapter 2), since a proper understanding of this is required before one can assess the quality of the measurements. A brief discussion (Chapter 3) is then given of issues relating to measurement precision, since these have a bearing on the data quality discussion as well as on interpreting the variability which exists in road data.

Assembling the databases for analysis was a major task, and this is explained in detail in the main body of the report (Chapter 4), as well as through supplementary information in Appendix C. The project has produced two CDs of data which are available (from the author) to other researchers interested in further research in this area.

The trends in deterioration for the various attributes are discussed in Chapter 5. Marked differences are observed in the variability in roughness data between State Highways (SH) and Local Authorities (LA), which is indicative of different levels of quality assurance. Chapter 6 describes the statistical analysis of roughness data, which was the only attribute considered suitable for analysis. The report closes in Chapter 7 with conclusions and Chapter 8 with recommendations for further work.

---

[1] RAMM Road Assessment and Maintenance Management System.

2. Measurement & Interpretation of Road Survey Data

2.1 Introduction

Regular data collection is essential for the proper monitoring of road condition, and thus the asset value. Accordingly, New Zealand road controlling authorities (RCAs) have annual data collection programmes. For these, data are collected visually and with automated equipment, such as roughness meters.

- **Manual Data.** This is a visual assessment of the pavement condition collected in accordance with the RAMM Rating Guide (Transfund 1997). The pavement distresses are recorded along a "Rating Length".
- **Automated Data.** Roughness is collected either using a laser profilometer or a response-type meter (e.g. NAASRA meter). State Highways are only measured with profilometers, while response-type meters or profilometers are used for local authority roads. Rut depths are collected with lasers or ultrasonics. Texture is collected with lasers, although mainly on State Highways.

The data are stored by all RCAs in their RAMM database.

This chapter describes how these data are measured. It provides background information to the data in the project databases. For completeness, the discussion includes the full range of data in the databases, even data not analysed as part of this project.

2.2 RAMM Pavement Distresses

2.2.1 Introduction

The definitions of RAMM distresses are from Transfund (1997) and NZIHT (2000). Since the study was looking at attributes which were likely to change over time, the data consisted of:

- Alligator Cracking,
- Longitudinal and Transverse Cracking,
- Rut Depth,
- Shoving,
- Potholes,
- Scabbing,
- Flushing,
- Edge Break.

The other distresses (joint cracks, broken or blocked channels, ineffective shoulders, surface-water channel problems, and inadequate drainage) were not included in the analysis since their deterioration is not suitable for time series modelling. The following sections describe how the eight distresses are measured in RAMM.

2.2.2 Alligator Cracking

Alligator (fatigue) cracking in RAMM is measured as the length (m) of individual wheel path, where the surface exhibits alligator cracking. Fine cracking confined to an area within 150 mm of the edge of the seal is not to be recorded as alligator cracking, as it is usually not caused by fatigue. As shown in Figure 2.1, a typical 50 m two-lane rating section has four wheelpaths so the maximum length of alligator cracking in the wheelpath (LWC) is 200 m.[2]

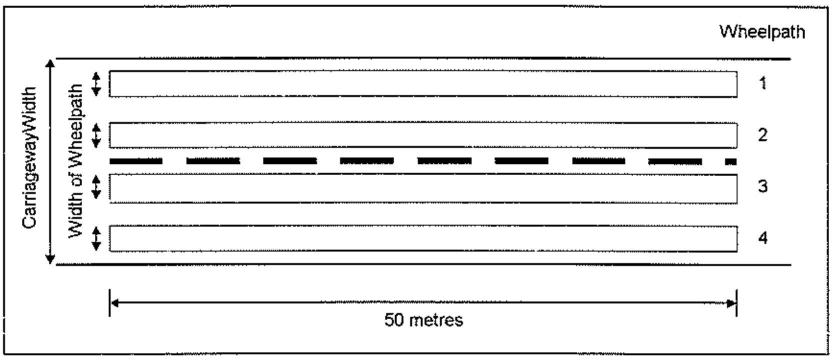


Figure 2.1 Wheelpaths and rating sections.

2.2.3 Longitudinal and Transverse Cracking

Longitudinal and transverse cracking (L&T) are cracks which appear along and across the carriageway. Large rectangular cracks are included with these as they are simply a more severe form of L&T which have formed a network. It is measured as the total length in m.

2.2.4 Rut Depth

As illustrated in Figure 2.2, rutting is defined in RAMM as the length (m) of individual wheel path where rutting (wheel tracking) exceeds 30 mm in depth (20 mm on SH maintenance group 1 roads), measured from a 2-m straight-edge laid transversely across the wheelpath. Only the length exceeding 30 mm is measured. Since there are 4 x 50 m lengths over a rating section, the maximum possible value is 200 m for this measure.

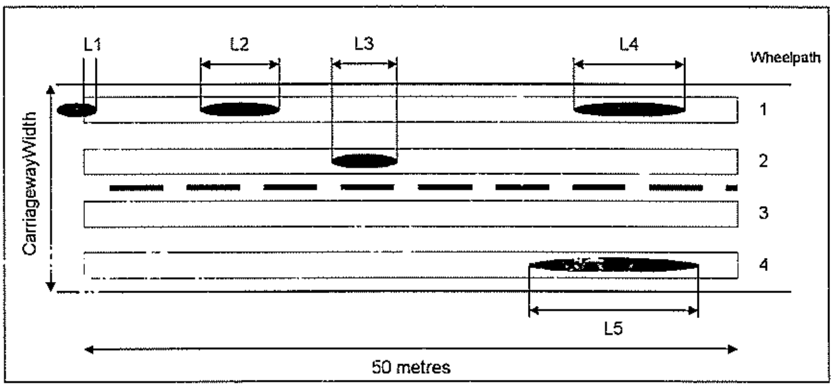


Figure 2.2 RAMM rut depth rating.

2 Since flushing, rutting and alligator cracking are usually found in the two wheelpaths, the total recorded must not exceed the number of traffic lanes x 2 x 50 m.

With the implementation of predictive modelling for pavement deterioration, there has been a shift of emphasis away from the RAMM approach of the length of pavement with rut depths greater than 30 mm, to the use of the mean rut depth. This trend is likely to continue as it is consistent with the output from the predictive modelling.

2.2.5 Shoving

As shown in Figure 2.3, shoving occurs in the transverse profile when material is displaced to form a bulge or heave alongside a depressed area. It is usually indicative of shallow shear failure.

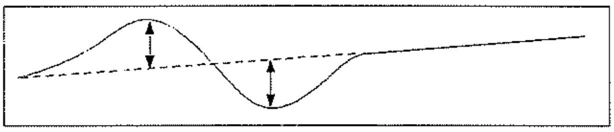


Figure 2.3 Shoving.

The length in m is recorded. If other faults occur within the area that is affected by shoving, e.g. if both alligator cracking and potholes appear in a shoved area, only the shoving is recorded.

2.2.6 Potholes

In the RAMM rating system a pothole is where the surfacing has broken to the extent that the layer (usually aggregate) below the surfacing is exposed. The break must have a maximum dimension of 70 mm or more to be rated as a pothole. Potholes are recorded as the number in the section.

2.2.7 Scabbing

Scabbing is a performance-based defect which occurs over time and under traffic when sealing chips have become separated from the bitumen in a chipseal. In an asphaltic concrete pavement, the aggregate loss from the mix is called ravelling (surface attrition), and is rated as scabbing.

A section is rated as scabbed when the area of carriageway (in m^2) within the inspection length that is scabbed is more than 10% of the surface area.

Sometimes “stripping” (Figure 2.4) is classified by RAMM condition raters (or raters) as scabbing. It differs from scabbing insofar as it is a construction defect which is visible almost immediately. Similar defects but with higher severities may be considered by some raters as potholes. Similar problems could be faced when rating delamination in multi-layer asphalt surfacing.

2.2.8 Flushing

Flushing occurs when the bitumen has risen so that the surface aggregate is just protruding (about 2 mm on grade 3 and 4 chipseals) or where the binder has risen to be level with, or over the top of, the surface aggregate. It can also occur because of punching of the chips into the existing seal.

Flushed surfaces have a lack of surface texture and a shiny or slick appearance. It is recorded as the length (in m) of wheelpath flushed.

Figure 2.4 Stripping of seal layer.

2.2.9 Edge Break

As shown in Figure 2.5, edge break is rated as such where the width of the sealed surface (including the sealed shoulder) is reduced by 100 mm or more from the nominal sealed edge. The length recorded shall be the length measured from the start of the taper leading up to the +100 mm edge break, to the point where the broken edge rejoins the line of the nominal seal edge.

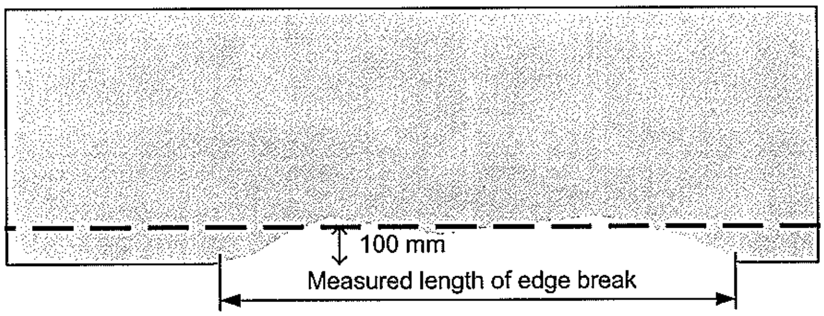


Figure 2.5 Example of edge break.

2.3 High Speed Data Collection

High speed data measurements (HSDA) are comprised of:

- Roughness,
- Rut Depth,
- Texture.

Each of these measurements is described below.

2.3.1 Roughness

Paterson & Scullion (1990) define road roughness as *"the deviations of a pavement surface from a true planar surface with characteristics dimensions that affect vehicle dynamics, ride quality dynamic pavement loads, and pavement drainage"*. To put it more simply, roughness is the "bumpiness" that one feels as one travels down a road.

In New Zealand two methods are used for measuring roughness: a response-type road roughness meter (RTRRM) and a profilometer. Examples of RTRRMs are the NAASRA and ROMDAS Bump Integrators. Examples of profilometers are the ARRB, WDM and PMS profilometers.

Figure 2.6 is an example of the NAASRA meter. Mounted on the rear floor of the vehicle, it records the relative motion of the vehicle floor to the axle. The measurements are expressed in counts/km.

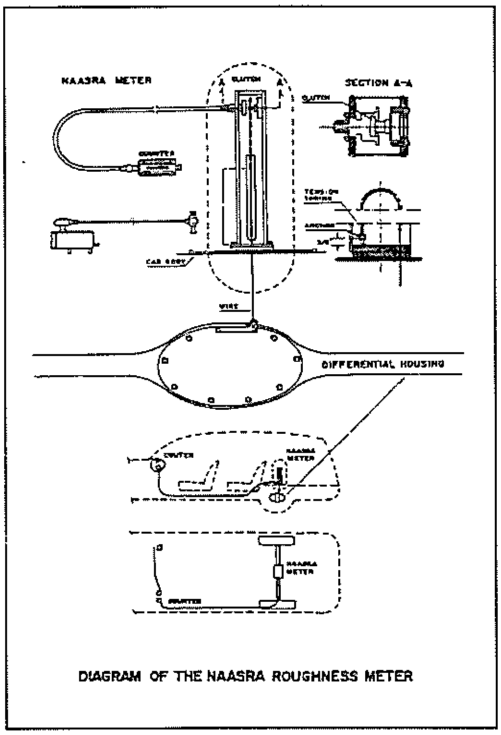


Figure 2.6 NAASRA Meter.

Figure 2.7 is an example of the ARRB TR Two Laser Profilometer. It uses a laser in each wheelpath to record the elevation of the chassis above the ground. Accelerometers are double integrated to establish the motion of the chassis through space. The difference between the movement of the vehicle through space and the height above the road is the elevation profile of the road. This elevation profile is then processed (often through post-processing) to calculate the roughness statistic IRI (International Roughness Index).

Figure 2.7 ARRB TR Two Laser Profilometer.

The underlying IRI model is a series of differential equations which relate the motions of a simulated quarter-car to the road profile. The IRI is the accumulation of the motion between the sprung and unsprung masses in the quarter-car model, normalised by the length of the profile. Mathematically this is expressed as:

$$IRI=\frac{1}{L}\int_{0}^{L/V}\left|\dot{z}_s-\dot{z}_u\right|dt \qquad (1)$$

where: IRI is the roughness in IRI m/km
L is the length of the profile in km
V is the simulated speed (80 km/h)
$\dot{z}_s$ is the time derivative of the height of the sprung mass
$\dot{z}_u$ is the time derivative of the height of the unsprung mass

Profilometers sample at a much higher rate than the NAASRA meter and so the roughness is often expressed on short sampling intervals, as low as 10 m. With the NAASRA meter it is more common to express the roughness on a 100-m basis.

The original NAASRA meter was in a "golden car" which was used only for calibration purposes. However, as described in the Transfund RAMM Manual (1997), calibration of NAASRA meters is now done against an ARRB TR laser profilometer. The conversion of NAASRA to IRI, and vice-versa, is done using the following conversion developed by ARRB TR:

$$IRI = (NAASRA + 1.27)/26.49 \qquad (2)$$

where IRI is the roughness in IRI m/km
NAASRA is the roughness in NAASRA counts/km

2.3.2 Rut Depth

Rutting is the permanent deformation of the pavement surface under traffic loading.

Automated measurements are made using equipment which can be called a "transverse profile logger" or TPL. These use lasers or ultrasonic transducers to

measure the transverse profile of a pavement as a vehicle travels over it at highway speeds. Figure 2.8 is an example of the ARRB TR multilaser profilometer.

Figure 2.8 ARRB TR Multilaser Profilometer.

Each profilometer has its own unique configuration for the positioning of the elevation sensors. Figure 2.9 shows the positioning for the ARRB TR multilaser profilometer where the sensors are positioned at different spacings. By comparison, the ROMDAS TPL has 30 sensors at 100-mm equal spacings.

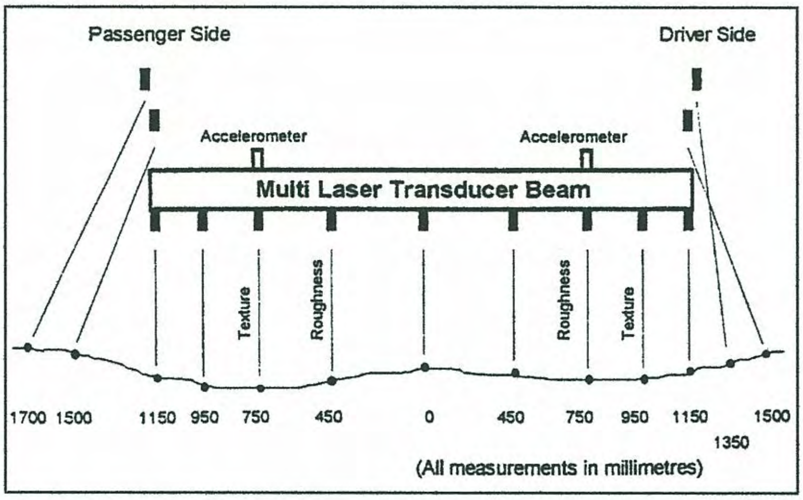


Figure 2.9 Positioning for ARRB TR Multilaser Profilometer lasers.

Irrespective of the technology used and the sensor spacing, the analytical approach is the same for all technologies. The elevations of each sensor result in the transverse profile being established, and the data are analysed to determine the rut depths.

Two basic models are used to do this: the **wire model** and the **straight-edge** model. The wire model is favoured since it is very fast in performing its calculations. It consists of stringing a simulated wire over the profile and then determine the rutting from the distance of this wire to the lowest elevation. A straight-edge model is more computationally demanding since it entails testing all the possibilities on the profile of placing the straight-edge.

Figure 2.10 is an example of the straight edge model. The straight-edge method has the advantage of being similar to the manual practices so is easier to verify. In New Zealand, the ARRB TR, WDM and ROMDAS TPLs all use the straight-edge, and the PMS system uses the wire model as its standard method.

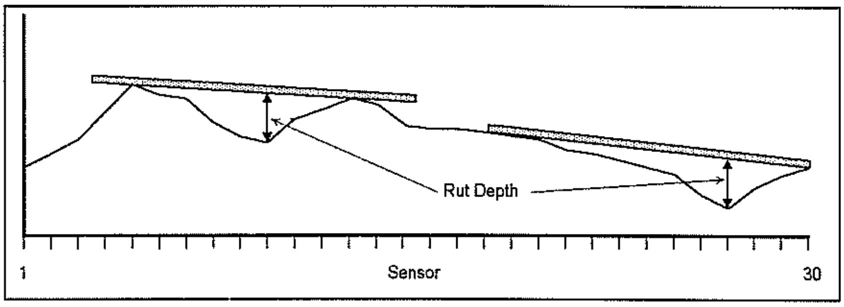


Figure 2.10 Example of Straight-Edge simulation.

One feature of profilometer measurements of rut depth that many users are unaware of, is that they **always** under-estimate the true rut depth. The reason for this can be readily visualised from the straight-edge simulation example shown in Figure 2.10. For the measured rut depth to correspond to the actual rut depth, the sensors would need to record the high and low points in each wheelpath. Since they are spaced at discrete intervals across the road, this is impossible.

Bennett (1998) tested the implications of discrete sampling of rut depth. The results are presented in Figure 2.11. The data were calculated by taking continuous transverse profiles and then calculating the rut depth as if the profile had been sampled at 100-mm intervals instead. The data clearly show the bias introduced from having discrete samples over the continuous sample.

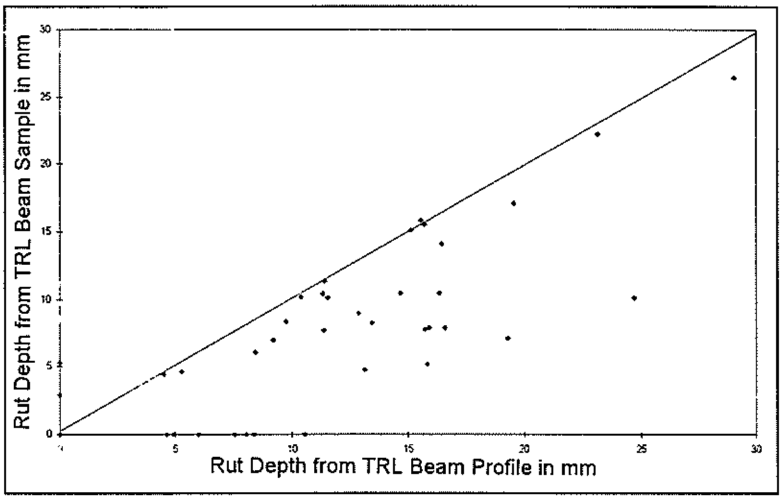


Figure 2.11 Effect of sampling on rut depth from continuous samples.

The amount of the bias will depend principally upon the lateral spacing of the sensors. The more sensors there are, and the closer they are together, the closer the readings will be to the true rut depth. However, it must be emphasised that a TPL will never give the same rut depth as that recorded manually.

2.3.3 Texture

The individual asperities or aggregate particles in a road surface constitute the macrotexture, and they range from about 6 to 20 mm in size. Macrotexture is therefore associated with the coarseness of the road surface that affects water drainage from the tyre print, tyre tread rubber deformation, skid resistance at high speed, and the friction-speed gradient.

Microtexture is the degree of roughness of the surface of individual aggregate particles exposed at the road surface, and has an amplitude that ranges typically from 10 to 100 microns. It is known to be a function of aggregate particle mineralogy and petrology, and is affected by climate/weather effects and traffic action. Also under this classification must be included the texture of bituminous and cement mortars, which may occupy major portions of the surface of asphalt mix and cement concrete surfacings between any exposed coarse aggregate particles.

The microtexture of the road surface affects the level of skid resistance at all speeds for dry and wet conditions. Surfaces with sharp microtexture projections have a high wet road skid resistance at low speeds but, without macrotexture, show a steep decline in friction as speed rises. Sharp microtexture projections are, however, associated with a high rate of tyre wear, and consequently the action of traffic polishes the surface, reducing its microtexture.

Figure 2.12 illustrates the principles by which contactless sensors work. A laser is used in conjunction with a photosensitive receiver to measure the variation in height from the sensor to the road. As the sensor moves over the road surface, the laser light is reflected from different points on the road surface so that different heights (D1 and D2) are detected by different diodes (d1 and d2) in the receiving array. The displacement in mm represented by the movement of the light from one diode to another is known, and so the variance in diode number can be converted into the root-mean texture depth in mm.

Based upon the angle of the light the **macrotexture** is measured. These systems cannot measure microtexture. Correlations have been made between the contactless sensors and the sand patch test. In New Zealand, BECA (1996) presented the equation:

$$TD = 2.1964 \ TDLASER \qquad (3)$$

where: TD is the texture depth (in mm) using the sand circle method;
TDLASER is the sensor measured texture depth (in mm) from the ARRB laser profilometer.

The texture data in the RAMM database was expressed in two units: MPD and IFI. These units are defined as:

- **MPD – Mean Profile Depth**. The draft ISO standard, ISO/DIS 13473 entitled "Characterization of Pavement Texture Utilizing Surface Profiles, Determination of Mean Profile Depth" defines a standard method for characterising surface texture. The MPD is calculated as shown in Figure 2.13.

- **IFI – International Friction Index.** The IFI was the outcome of a major experiment by PIARC in 1992 to harmonise friction measurements. The IFI is composed of two numbers: the calibrated wet friction at 60 km/h (F60) and the speed constant of wet pavement friction (S_p). It is calculated from a measurement of pavement macrotexture and wet pavement friction.

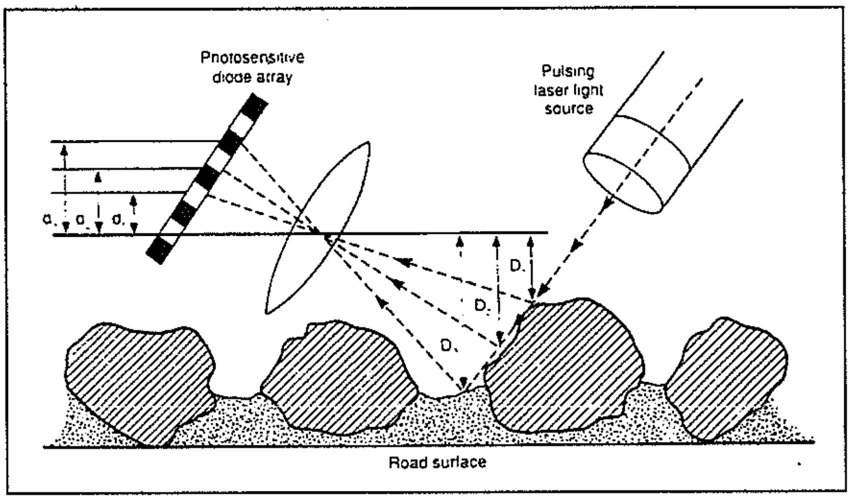


Figure 2.12 Texture measurements (Millard 1993).

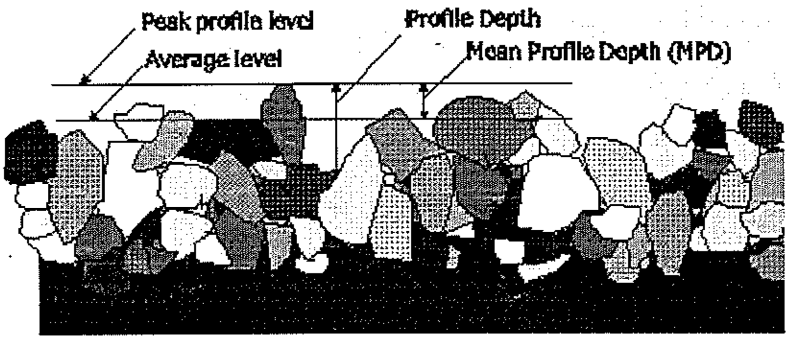


Figure 2.13 Calculating mean profile depth.

2.4 Sources of Data Errors in Road Surveys

The annual survey data represent a "time" series, where the spatial location and order of the sampling points play the role of time. The goal of the project is to develop a statistical approach, and a simple method for applying it, to evaluate these data for consistency.

There are three sources for errors in data collection:

- random errors,
- systematic errors, and/or
- operator errors.

If the current data do not compare well with the data from the previous surveys, this is due to one (or more) of four reasons:

- the current survey is wrong,
- the previous survey(s) were wrong,
- the data were collected or referenced to the wrong location, and/or
- the pavement has been modified.

Ideally, if all data were correctly referenced and accurately collected there would be a definite trend, such as illustrated in Figure 2.14A. However, this is seldom if ever the case and there is almost always scatter in the data, as in Figure 2.14B. The complicated issue is to decide whether or not the scatter is caused by systematic measurement errors or by location referencing errors (i.e. operator errors). If they are caused by random measurement errors, the data cannot be adjusted. If they are caused by location referencing errors, one may be able to get a good trend as shown in Figure 2.14A by possibly adjusting the chainage of the data forward and/or backward.

Pavement maintenance further compounds things since it leads to discontinuities which deserve to be in the data. This is shown in Figure 2.14C where a shape correction leads to a significant reduction in the roughness. Although these points have the same values as in Figure 2.14B, to fit a trend line to the data as suggested in Figure 2.14B would be inappropriate. Instead, the analysis would need to take account of the shift in condition.

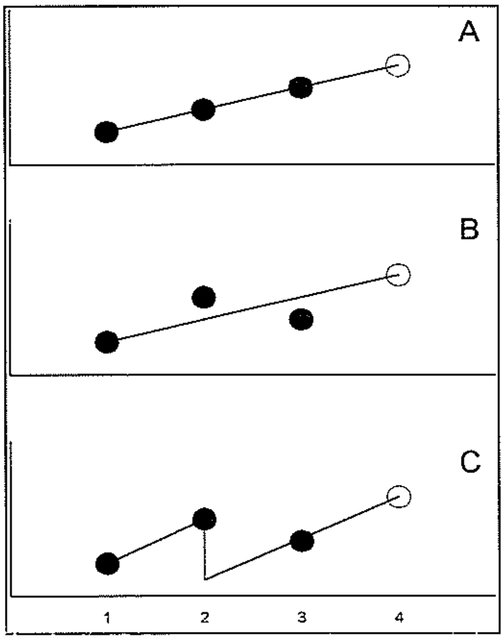


Figure 2.14 Hypothetical data from four surveys.

It should be noted that this problem is not confined to New Zealand, but is present throughout the world. Yet the issue of systematic techniques for evaluating data quality does not appear to have been considered in the technical literature

3. Implications of Measurement Precision

3.1 Introduction

In considering the quality of road survey data one needs to assess both the accuracy and repeatability of the measurement.

Accuracy is the ability of the instrument to predict the reference roughness without bias. This is expressed in terms of NAASRA counts/km or, for State Highways since 1996, also in terms of IRI m/km.

Repeatability is the ability of the instrument to produce the same result in multiple runs with minimal random error. It is important that an instrument be repeatable since a lack of repeatability suggests a random error source during the measurement. Karamihas et al. (1999), in discussing profilometers, give a number of factors which affect the repeatability of measurements including:

- Surface shape,
- Temperature variations, particularly with Portland Cement Concrete pavements,
- Seasonal variations which affect the volume of the subsurface layers,
- Transverse variations in the roughness,
- Pavement distresses such as cracking and rutting,
- Lateral positioning of the vehicle during measurements,
- Profile driver and operation.

With response-type meters such as the NAASRA meter, the main factors affecting repeatability are the vehicle operation, lateral positioning, and driver behaviour.

Assuming that the vehicle's measurements are accurate, i.e. it is properly calibrated, the repeatability is the key factor affecting the precision of a measurement.

3.2 Implications of Precision on Deterioration Trends

To illustrate the implications of repeatability on the observed pavement deterioration, consider Figure 3.1 which shows six situations drawn from the databases. It shows the mean measurement and hypothetical error bars representing the confidence intervals around the measurement. On the basis of the mean, six different cases are presented for the trend in pavement deterioration, shown by the broken lines in the middle of the figure.

- **Case A:** Slight increase
- **Case B:** Slight decrease
- **Case C:** Large increase
- **Case D:** Large decrease

- **Case E:** Major increase
- **Case F:** Major decrease

The bottom of the figure shows the possible deterioration associated with the confidence intervals, assuming that pavements do not improve over time.

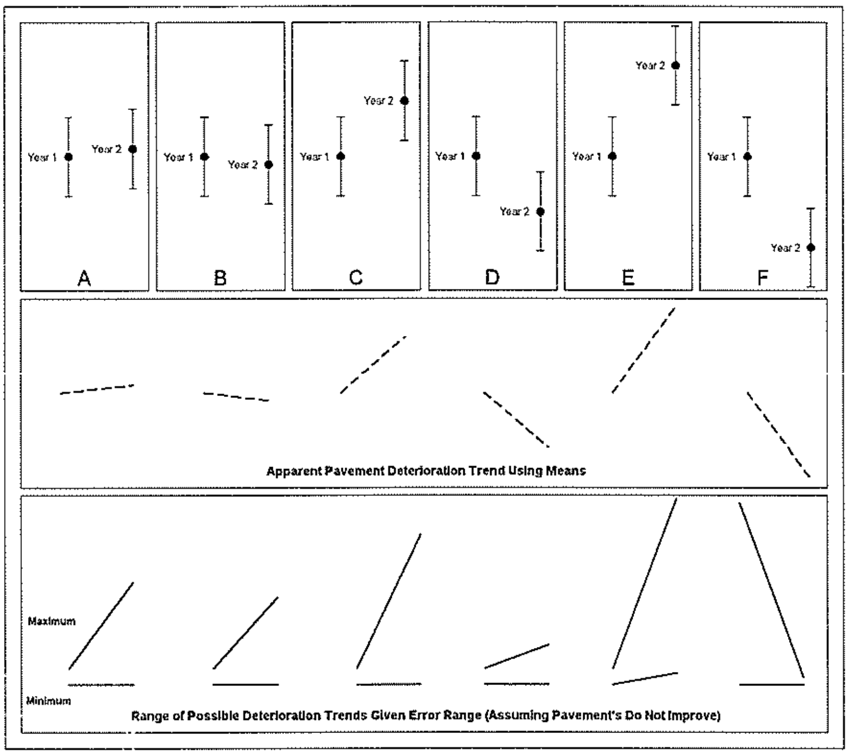


Figure 3.1 Implications of confidence intervals on observed deterioration.

It is apparent from the examples in Figure 3.1 that the only way to obtain a reliable measure of pavement condition is to have measurements made as precisely as is practical. This serves to decrease the size of the confidence interval, thereby allowing trends to be clearly observed. A lack of precision means that it is not possible to discern trends in pavement condition, nor is it possible to fully assess the quality, or lack thereof, of the survey data. This issue will be explored later.

3.3 Precision of Measurements

No data were available from which to investigate the precision of visual condition measurements. Data were available for roughness measurements, and Appendix A describes the outcome of an analysis of this data.

The conclusion was that profilometers have a standard deviation on the order of 2%; and a properly calibrated response-type roughness meter of 3%. While one cannot translate these precisions directly to vehicles operating in a survey, where there is a single pass once a year, it does show that under controlled conditions the instruments can have significant variability.

This variability has an important implication when it comes to investigating pavement deterioration trends since, based on the work in Paterson (1987), roughness typically progresses at less than 5% per year. With the measurement precision on the order of the rate of change in pavement condition, one must have data from over several years to obtain an indication of the pavement deterioration rate. This highlights the importance of having precise measurements of pavement condition when trying to investigate pavement deterioration.

4. Assembly of Project Databases

4.1 Introduction

For this project data were obtained from the following RCAs:

- Napier State Highway,
- Northland State Highway,
- Auckland City Council,
- North Shore City Council,
- Southland District Council, and
- New Plymouth District Council.

No checks were made or specific criteria used for selecting the data. It represented "typical" values as provided to the RCA from the data collection consultant. This is the data that has been used for making decisions on the road network.[3]

The assembly of the databases proved to be much more difficult than anticipated and was an extremely complicated and time-consuming process. This was due to a variety of factors, one of the principle ones being the architecture of RAMM and its lack of proper location-referencing controls which necessitated the use of dynamic segmentation to establish the necessary analysis sections.

Appendix B describes the entire process by which the databases were created. Should other researchers be interested in the data, two CDs are available from the author which contain the sample data as well as all intermediate steps, templates as well as other material. Appendix C describes the contents of these CDs. The steps involved in establishing the databases are described in the following sections.

4.2 Assembly of Databases

The databases were created using the HDM-4 Information Management System (HIMS), a road management system application developed by HTC before this project commenced.

The data were provided as ASCII text files, created by using an extraction routine from RAMM[4]. The following source tables were used to obtain the data:

carr_way
csurface*
topsurf
pavelayr*

[3] Some minor corrections were made to the data to ensure consistency. For example, one database used L/R for roughness in some years; L1/R1 in others. They were standardised to L1/R1.

[4] The best source data format for organising data is *.unl format with the road.sql definition.

p_struct (paylayr and p_struct can give the pavement table)
hsdrough
hsdrutt
hsdtext
hsgeomtry
rating
skidres
rough
roadname
traffic*
treatlen*
treat*
road.sql*

*: source data may not be directly used

Because of doubts as to the validity of very old data, only data from 1990 were included in the analysis. The steps taken to organise data into the final structure for analysis were as follows:

- The source data were checked to ensure data values were consistent, e.g. changing all "Left Lane 1", or "L" to "L1" etc. The lack of consistency between the RCAs, and even between years of data collection for the same RCA, was a major deficiency in the RAMM databases.

- Carriage width data were transformed into a surface and pavement data table. Having done this, the surface and pavement data were organised into different year/lane. The data were separated into different lanes according to the formulae below (for two lanes, L1 and R1):
  Left lane: Cway_width/2 – offset >2
  Right lane: Width+ offset – cway_width/2 >2

  Only treatments more than 2 m wide were considered. The length of treatment was automatically solved when data were transformed. For pavement data only the first layer (layer_No=1) was considered.

- Other data were organised into different year/lane. This gave a result expressed as: Roughness_1998_L1, Roughness_1999_L1 etc. Since a road may have more than one left lane or right lane, only data for the farthest left lane or the farthest right lane were used.

- The surface and pavement data were then transformed into data for a specific year/lane. Since a survey may have occurred before or after a treatment, it was necessary to transform the surface and pavement data from the same year and the previous year into another data table of the same lane. Later, by comparing date of survey and treatment, the treatment was assigned to a date before or after the survey. For example:
  Transform Surface_1998_L1 into Roughness_1998_L1
  Transform Surface_1999_L1 into Roughness_1999_L1

- New 100-m sections for all roads were created. This was done to counteract the problem of data in RAMM not being well referenced, making data from different years not directly comparable. The 100-m sectioning was just a generic rule adopted for this analysis; longer lengths could have been used. Since roads usually do not end in a multiple of 100 m, the right value was used (and very small sections were joined). It should be noted that the State Highway data were stored in 10-m intervals so the aggregation to 100-m results in a loss of resolution. However, it also served to compensate for any small location-referencing problems which may have influenced the results.

- All data except surface and pavement, whose information were already included, were transformed into the new 100-m sectioning system.

- The data from different years for the same lane were organised into one table. For example, NAASRA of left lane 1 were organised into the table Rougness_L1, which included NAASRA from different years for that lane.

- Checks were made to ascertain whether a treatment was done in a given year, which would affect the road condition. If a treatment was done before a survey, it was considered that the treatment was done in that year since the survey would reflect the changed pavement condition. If a treatment was done after the survey, then the treatment was assigned to the next year.

- Data were then exported into its final format as described in Section 4.4.

4.3 Treatments

As mentioned earlier, treatments will influence the condition of a pavement. For example, a reduction in roughness may be due to the fact that the pavement received an overlay or shape correction.

Originally, it was intended that a treatment hierarchy would be used which would reflect the magnitude of the treatment. For example, patching would be anticipated to have relatively little impact on the 100-m roughness, whereas shape correction would have a major impact. However, it was found that the data required to have such an approach was not readily available in RAMM without resorting to some detailed maintenance cost analysis which was beyond the scope of the project.

Consequently, the databases had only two values which reflected the type of maintenance applied to the pavement[5]:

3 = Resurfacing
4 = Overlay/Rehabilitation/Shape Correction

In all years when a treatment was not performed the flag was set to 0. As described in Section 4.2, if a treatment was performed in the year before the data collection

[5] The original hierarchy had 1 = Minor patching, pothole filling, *etc.*; 2 = Major patching; 3 = Resurfacing; 4 = Overlay/Rehabilitation/Shape correction.

survey was executed that year was given a flag of 3 or 4 (as appropriate). If, however, the treatment was applied after the data collection survey, the flag was instead inserted into the data for the following year.

4.4 Database Structures

The database structures were essentially the same irrespective of the distress of interest. There were the following fields:

- **Road_ID**. RAMM road identification number.
- **Road_Name**. RAMM road name.
- **Start_M**. The start chainage in m where the data recording began.
- **End_M**. The end chainage in m where the data recording ended. The data applied over the interval Start_M to End_M.
- **Lane**. The lane the data applied to. This followed the standard RAMM convention of L1/R1. Some local authorities had "B" which meant that the provider collected the data in one direction, but applied the same value to both directions. Others had L/R. As described earlier, all were standardised to L1/R1.
- **Data.** The first five fields defined the location of the data. These were followed by fields indicating the year of the data and, where appropriate, the wheelpath where it was collected. A typical example for rating rutting consisted of:
  - 1992_Rutting
  - 1993_Rutting
  - 1994_Rutting
  - 1995_Rutting
  - 1996_Rutting
  - 1997_Rutting
  - 1998_Rutting
  - 1999_Rutting
  - 2000_Rutting

  There were some deviations, for example, the wheelpath IRI consisted of the following fields for State Highways. The first 5 fields were the left wheelpath roughness for 1996-2000; the second 5 fields the right wheelpath roughnesses for 1996-2000:
  - 1996-LWP-IRI
  - 1997-LWP-IRI
  - 1998-LWP-IRI
  - 1999-LWP-IRI
  - 2000-LWP-IRI
  - 1996-RWP-IRI
  - 1997-RWP-IRI
  - 1998-RWP-IRI
  - 1999-RWP-IRI
  - 2000-RWP-IRI

- **Maintenance**. As described above, depending on the year, the maintenance treatment applied in the year different year was flagged as 3 for resurfacings; 4 for overlays/shape corrections. The maintenance fields were:
  - Maint_1996
  - Maint_1997
  - Maint_1998
  - Maint_1999
  - Maint_2000

4.5 Data for Analysis

Each database was comprised of a series of tables for analysis. The tables contained the data for discrete chainages in the form of a time series, with the readings for each year for the same chainage in different columns following the chainages. The files and their contents were as follows:

RAMM Rating Data	
Rating-Alligator	Alligator Cracking
Rating-Broken	Broken Channel
Rating-Edge-Break	Edge Break
Rating-Flushing	Flushing
Rating-High-Lip	High Lip
Rating-L&T	Longitudinal and Transverse Cracking
Rating-Pothole	Potholes
Rating-Rut	Rut Depth
Rating-Scabbing	Scabbing
Rating-Shoving	Shoving
High Speed Data	
Crossfall	Pavement crossfall
Curvature	Horizontal curvature
Gradient	Gradient
IFI-Texture	International Friction Index
Max-Rut-Depth	Maximum Rut Depth
Max-Texture	Maximum Texture
Mean-Rut-Depth	Mean Rut Depth
Min-Rut-Depth	Minimum Rut Depth
Min-Texture	Minimum Texture
MPD-Texture	Mean Profile Depth Texture
Roughness-HSD	Roughness in NAASRA and IRI
Roughness-NAASRA	NAASRA Roughness
Roughness-Wheelpath-IRI	Roughness in IRI m/km
Sdev-Rut-Depth	Standard Deviation of Rut Depth

Not all data were available for every data set. For example, few local authorities had high speed-profilometer data which meant that there was no texture or rutting. The roughness was limited to NAASRA roughness. By comparison, the State Highways had High Speed data for 4-5 years with all attributes.

In addition to the manipulated data, the databases also contained the raw ASCII text files exported from RAMM used as source data. These were listed based on their original table names from RAMM, and given the suffix #asc.

4.6 Database Issues

The preparation of the databases proved to be a very complicated process. The following are several issues that arose as a result of this exercise.

4.6.1 Location Referencing

With some roughness data for local authorities, there was a problem with regard to the survey start and end chainages.

Unless the sections have changed, all road surveys should start at the same location and end at the same location. Thus, the length of the road is constant between years. However, this was not always the case and showed a need to improve both the data collection and data processing procedures.

Figure 4.1 illustrates this with data from three years of survey. If the data were correctly referenced from the start of the road, all the points for the three surveys would be at the same chainage. Instead, one was offset by 55 m (i.e. 27155-27255; 27255-27355, etc.), a second by 86 m (i.e. 27186-27286; 27286-27386, etc.), and the third (properly) every 100 m (i.e. 27100-27200; 27200-27300 etc.). These problems were not found with the State Highway data which appeared to have much better location-referencing controls.

These problems extended to the end chainage of the road. An example of this for one road had surveys in successive years ending at chainages 30986; 30966; 30986 and 30985. While part of the differences may have been related to the start chainage problem identified above, it arose even on sections where there were no start chainage problems. Thus, it is more indicative of poor data processing on the part of the data collection contractor, or lack of proper controls by the RAMM data manager.

4.6.2 Changes to Inspection Lengths

The visual RAMM data were collected over inspection lengths. Around 1998 there were changes to treatment lengths, and this impacted on the location of the inspection lengths. The year where this discontinuity arose depended upon the RCA. It also did not arise for every road section.

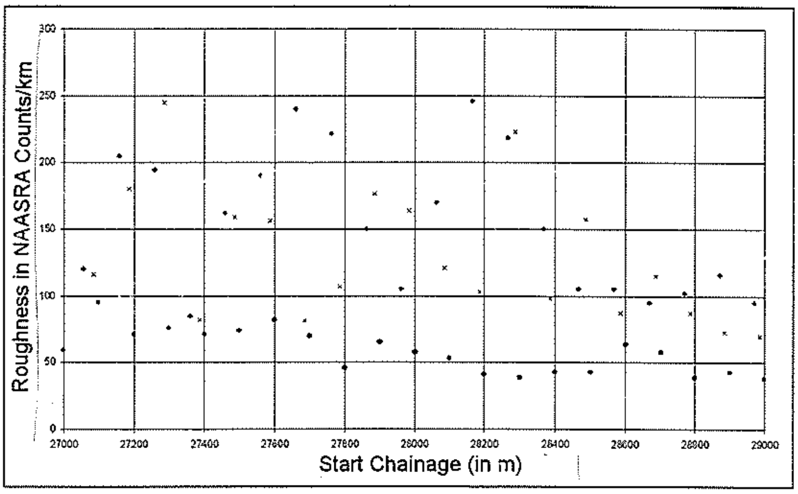


Figure 4.1 Example of referencing problem.

The process adopted was to extract the longest continuous time series of data from these files. In practice, this meant that if the rating section locations had changed, the data were usually based on the earliest data, and if it was unchanged it meant that data up to the present were used.

4.6.3 Lane Identification

With the roughness data, some data were found to have been referenced L/R by one contractor, while in the same database another contractor referenced them L1/R1 (which is more appropriate). In some databases "B" was used to reflect that the data were collected in both lanes.

4.6.4 Data Variability

The data were both consistent and yet variable. While some variability was due to the fact that the roads had been maintained between years, a preliminary review of the data suggests that there could be significant issues with regard to the quality of the road survey data in the RAMM databases.

5. Road Deterioration Trends

5.1 Introduction

As described by Paterson (1987), pavement deterioration is dependent upon time, traffic, the environment pavement strength and the surface condition. Between successive years it may be either imperceptible or quite marked.

Given the precision of roughness measurements described in Section 4, there will always be variability in the data for the same section between successive years. However, there should be discernible trends if the underlying data are reliable.

As a precursor to the statistical analysis, the road deterioration trends were investigated for two sets of data:

- Roughness, and
- RAMM visual rating.

5.2 Roughness Data

Appendix D shows examples of roughness trends in 100-m intervals for 500-m sections of State Highways and an RCA. The figure below is typical of the data for State Highways[6].

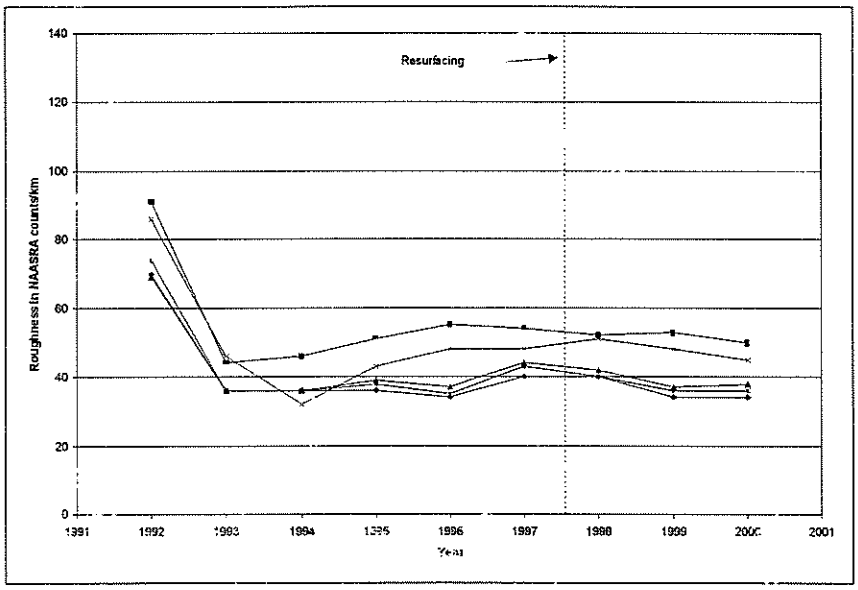


Figure 5.1 Example of State Highway Roughness progression.

[6] The State Highways had data recorded with the NAASRA meter before 1994; with an ARRB laser profilometer 1994-96; and with a WDM profilometer 1997-2000.

The data show relatively little variability and are generally consistent between years. There was a decrease in roughness after the resurfacing in 1997, although an earlier maintenance in 1992-93 was not in the database.

The local authority data tended to show much more variation than the State Highway data, which suggests a lower level of quality control. Figure 5.2 is typical of the RCA data. As with the State Highway data, there were situations where maintenance had obviously been done, but was not recorded in the RAMM database.

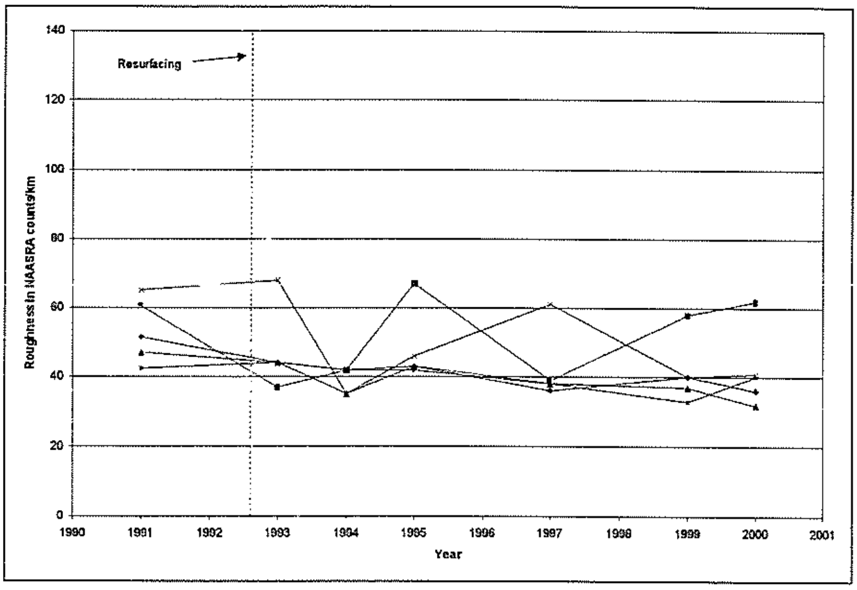


Figure 5.2 Example of RCA Roughness progression.

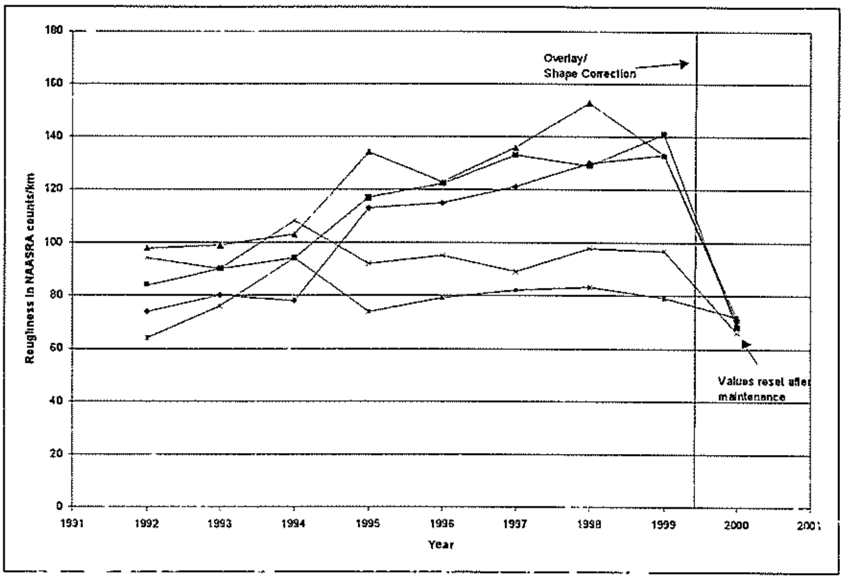


Figure 5.3 Example of effect of maintenance on Roughness progression.

The importance of maintenance on roughness progression cannot be over-emphasised. Minor maintenance, such as patching, may lead to small changes in the roughness, whereas resurfacing may lead to moderate changes, depending upon the maintenance activities done before resurfacing. Shape corrections and overlays lead to substantial changes in roughness, particularly since these are often triggered by high roughnesses. This is illustrated in Figure 5.3 which is from a State Highway where the data show a clear re-set in condition after maintenance in 2000.

Unfortunately, the maintenance records available in RAMM are deficient for monitoring whether or not maintenance has been performed. The maintenance activities were not always recorded in the database (as evidenced by the maintenance missed in 1992-93 in Figure 5.1), but only resurfacing and shape corrections/overlays were available. The failure to record other maintenance activities such as patching means that discontinuities in the deterioration trends which appear to be anomalies may in fact have been due to maintenance on the pavements.

Figure 5.4 shows a conceptual example of how maintenance complicates the analysis. In the top of the figure (A) is a series of observations on a pavement which is not deteriorating. After maintenance the data are also constant. However, the maintenance serves to shift the series, making trend analysis complicated. Were one to use the differences in the readings (B) the trend is obvious, with the data after the maintenance representing an outlier. Statistical techniques are available to account for this, but given the nature of the data they may be stretched.

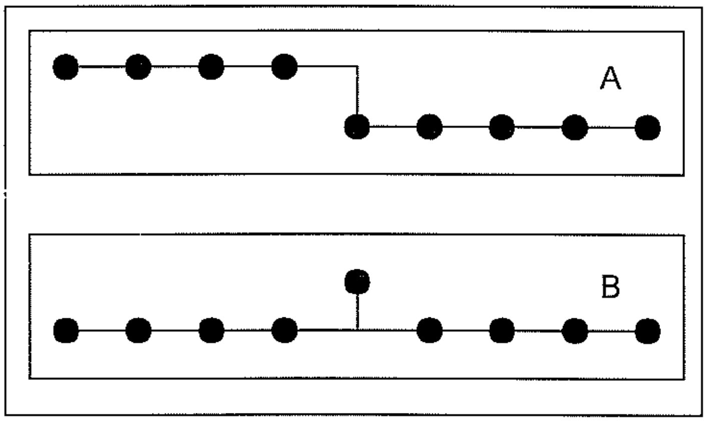


Figure 5.4 Example of effect of maintenance shift on trends.

5.3 RAMM Rating Data

Visual data which are subjectively recorded, often by different individuals in different years, would be expected to have less reliability than quantitative data such as roughness data. Furthermore, depending upon the sampling procedure, the locations where the surveys were made can conceivably change over time. Compounding this situation is the fact that many of the RAMM distresses would be addressed by patching and minor maintenance, which could not be easily extracted from the RAMM database. Thus only resurfacings and overlays/shape corrections were in the project databases (see Section 4.3 of this report).

To illustrate this, consider Figure 5.5 which shows the trends in alligator cracking for 3 x 100 m sections from a local authority database. For two sections the cracking was reset to 0 in 1998 because of maintenance, and stayed at that level thereafter. Unfortunately, this maintenance was not recorded in the RAMM database so there was no way of accounting for it in an analysis.

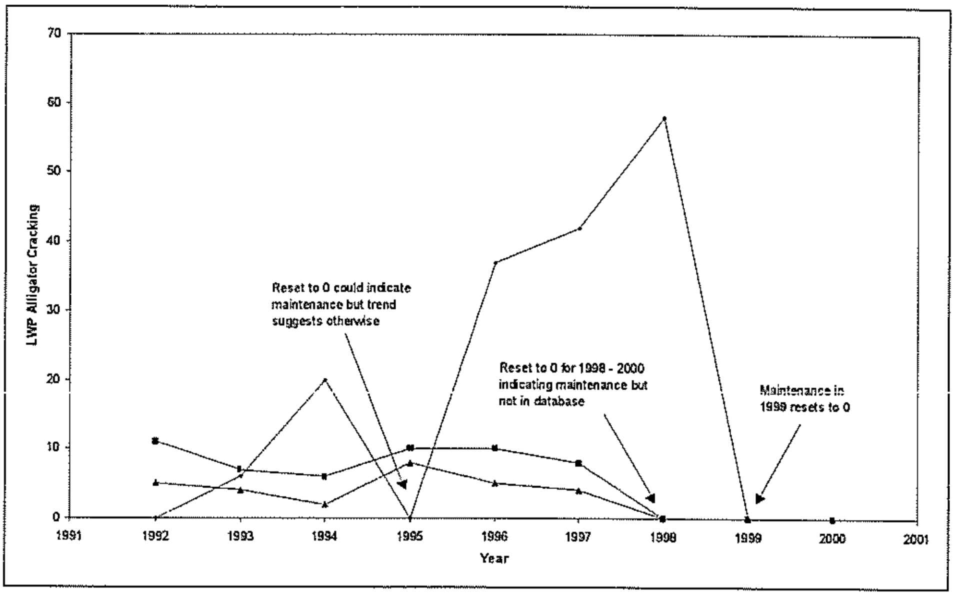


Figure 5.5 Example of Alligator Cracking time trend.

The third section (not adjacent to the other two) had maintenance recorded in 1999, and this corresponded to the cracking being reset to 0. However, a 0 level for cracking in 1995 looks to be incorrect given the general trend of the distress and its continuity from 1995 onwards. Although the other distresses were evaluated for this section, it proved impossible to confirm whether or not this 0 value in 1995 was in fact correct.

This situation was present with other distresses. For example, Figure 5.6 shows data on scabbing. One section, which had 150 m of scabbing in 1995, was given a surfacing treatment which is reflected in the values from 1996 onwards. However, the other two sections (which were from a different road) also showed a decrease in 1995, but no maintenance treatment for this activity was recorded in the database.

The issue of measurement precision appeared to be present with some of the data. An example of this is shown in Figure 5.7 which shows the RAMM rut depth rating for two adjacent 100-m sections. The same pattern is followed on both sections between years and, while this is possibly related to the effect of local maintenance, a more likely explanation is that the variation is related to the way in which the data were recorded. As with some other road sections, the decrease to 0 after 1997 was due to a maintenance activity which was not in the database. The same can be observed in Figure 5.8 which shows RAMM edge break over time. The resets to 0 in 1996 were not recorded in the database.

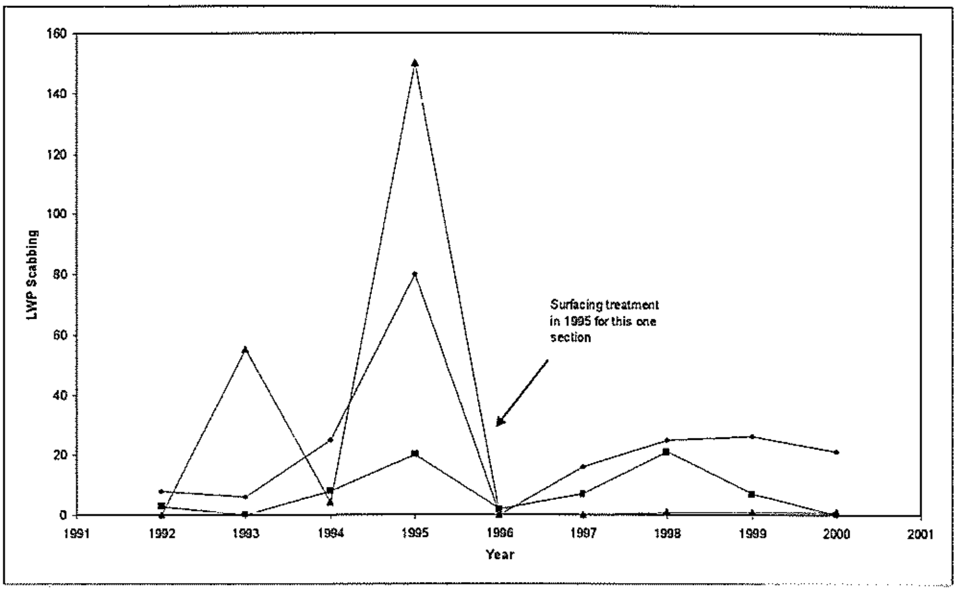


Figure 5.6 Example of Scabbing time trend.

Figure 5.7 Variation in RAMM Rut depth over time.

Some distresses, such as potholes, did not show any form of time trend. This would be anticipated since potholes are usually patched promptly to prevent them progressing and forming more major distresses. This also appeared to be the case with flushing and shoving which were generally reset to 0 very soon after appearing.

5.4 Implications of Data Trends on Statistical Analysis

The review of the data trends indicates that the statistical assessment of data quality will be problematic because of the impacts of pavement maintenance.

The failure for major maintenance (resurfacings, shape corrections) to be reliably recorded in RAMM, or for minor activities to be adequately recorded for the purposes of time series analyses[7], means that "noise" will be introduced to the data which are related to unique, external factors. These lead to discontinuities in the deterioration trends, which make any form of trend analysis difficult at best, and most likely meaningless in many situations since we are compounding the other problems such as lack of precision and small sample sizes.

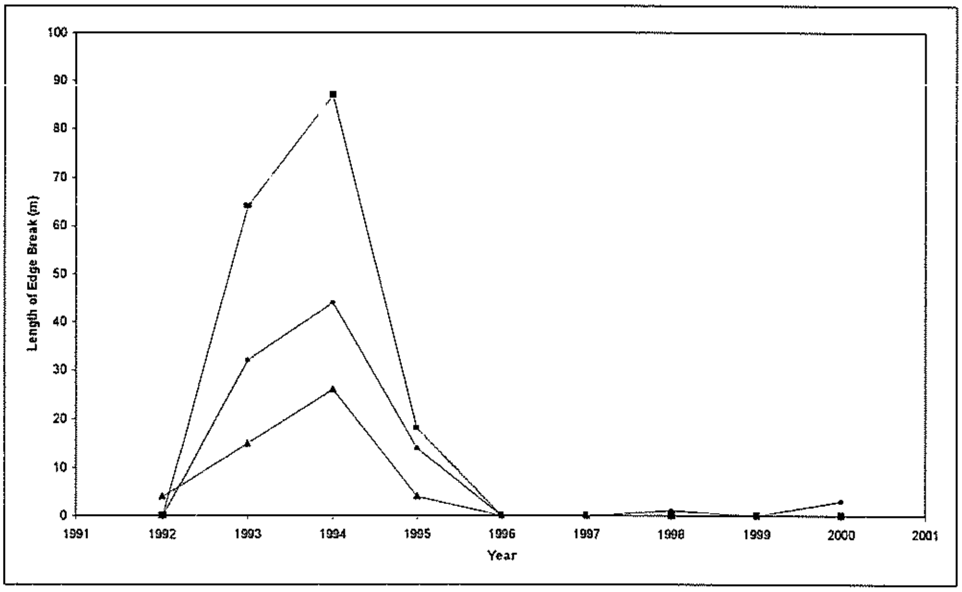


Figure 5.8 Variation in RAMM Edge Break over time.

The data can effectively be divided into three discrete groups:

- **No maintenance**. These sections would be suited to trend analysis. However, when dealing with time series data over 7 years, it is very rare in New Zealand to have data with no maintenance because of the practice of regular resurfacings. Those sections without maintenance data may in fact be reflecting a failure to record maintenance instead of no maintenance being performed.
- **Maintenance during the time series**. This is the most common and problematic situation. The recorded (or unrecorded) maintenance results in a discontinuity in the deterioration trend. Either statistical techniques for identifying and eliminating discontinuities due to maintenance will have to be developed, or the data will need to be filtered so that the data prior to the maintenance are not included in the analysis.

---

7 As described in Chapter 4 of this report, the information for minor maintenance could be accessible through the maintenance cost data, but this proved to be too complicated to extract for the purposes of this project.

- **Maintenance in the last year**. If the maintenance was performed in the last year it is possible to use the data from previous years to establish a trend in deterioration, although this trend may be altered due to the maintenance activity. These situations are useful for testing the statistical technique for identifying anomalous data.

With regard to the RAMM rating data, this is the most heavily influenced by the problem with identifying maintenance, since many of the distresses will be reset by minor maintenance which is not in the database. Thus, until it is possible to accurately identify which sections have received minor maintenance activities, there is no use in undertaking statistical time series analyses on RAMM data.

The roughness data offers greater scope for analyses, so the focus of the work shifted to investigating the statistical analysis of roughness data. This is described in Chapter 6.

6. Statistical Analysis of Roughness Data Quality

6.1 Introduction

As described in Chapter 5, the RAMM data were considered unsuitable for time series analyses because of the impact of maintenance which was not recorded in the RAMM database. The analysis therefore focused on roughness data which would not be as influenced by minor maintenance as the RAMM data. In addition, since roughness is measured quantitatively compared with the qualitative nature of RAMM data, the roughness data is the most reliable data available.

6.2 Processes and Process Variability

As shown in Appendix A, when measuring the roughness on the same section of road the vehicle may record 61, 63, 61 and 59 NAASRA counts/km in repeat runs. All processes have this type of variability.

The processes can be partitioned into two components.

Natural process variation, also called "common cause" or "system" variation, is the naturally occurring variation inherent in all processes. In the example above, the roughness varies around the true roughness which, if the mean is representative of this value, would be 61 NAASRA counts/km.

Special variation arises an extraordinary event occurs to the system. For example, if during one of the runs the roughness vehicle was forced to brake, and this could give a reading of 70 NAASRA counts/km.

In the context of road data, where the conditions change over time, it is necessary to have a technique which accounts for this, instead of assuming that the process is constant.

6.3 The Sample Size Problem

The roughness data analysis immediately highlighted a major limitation of the proposed time series analysis approach: the small sample size. At best, there were 9-10 years of data available, at worst there was less than this due to roughness not being recorded in certain years. These are very small, many would say inadequate, sample sizes for time series analysis since the modelling processes reduce the number of degrees of freedom. While statistical techniques exist for estimating missing time series values, they would not be appropriate here due to the sample size limitations.

Box et al. (1994) do not address this issue of sample sizes, but the general rule is that the larger the sample size the more reliable the estimation. When dealing with small samples the key is the "signal-to-noise ratio" which is the ratio of the explained to unexplained variance. If you have a small sample but a high signal-to-noise ratio, then one can trust the identification process. However, if there is a large amount of error due to outliers, or changes in parameters, or changes in variance then even with large sample size it may not be possible to properly identify the model form.

One method for overcoming the sample size limitation would be to look at the spatial component of the measurements. There is likely to be some correlation along the road with the condition and this additional information could be used in conjunction with other outlier techniques for identifying the quality of the data. Unfortunately, this approach was beyond the scope of this project.

6.4 Box-Jenkins Time Series Analyses

After considering the various options, the analysis approach adopted was based on univariate Box-Jenkins (B-J) time series modelling. This describes a time series of condition data as a function of its own past values. The purpose of the B-J process is to reduce the time series with underlying structure to "white noise"[8] which is the predictable portion of the time series. This can be used to forecast future values of the series and thereby identify the quality of new data against the previous trend. This meets the goal of the research project for identifying the overall quality of a measurement.

The software application Autobox[9] was adopted as the analysis tool. Specifically designed to apply the B-J process, it is a tool which also has a batch-mode option that allows for the analysis of large data sets such as were established for this project.

Autobox uses a three stage iterative process for its analysis of:

- **Identification**: A tentative model form is selected by examining a plot of the series and several key statistics;
- **Estimation and Diagnostic Checks**: The parameters in the identified model are estimated; the model's residuals are examined for model sufficiency and necessity; and
- **Forecasting**: The model is used to generate forecasts of the future value of the time series.

---

8 Box et al. (1994) define white noise as follows: "The stochastic models are based on the idea that a time series in which successive values are highly dependent can frequently be regarded as generated from a series of *independent* "shocks" a_t. These shocks are *random* drawings from a fixed distribution, usually assumed Normal and having a mean zero and variance σ^2_a. Such a sequence of random variables $a_t, a_{t-1}, a_{t-2} \ldots$ is called *white noise*".

9 Autobox is a commercial application specifically designed for time series analyses using Box-Jenkins. One of its features, which made it particularly attractive for analysing road data, was its ability to analyse large batches of data. Details on Autobox are available from www.autobox.com.

Road condition data usually form a non-stationary process, as evidenced by the change in the mean over time, or the non-uniform variance over time. It is therefore necessary to transform the data so that it is stationary and not trending. Autobox transforms the data using the differences between successive readings, i.e.

$$\nabla z_t = z_t - z_{t-1} \tag{4}$$

where: ∇z_t is the difference in readings between time t and time t-1,
z_t is the reading at time t.

While Box et al. (1994) use linear factors, Autobox have polynomial factors of the form $(1 - B^o)^d$, where o is the order of the differencing factor, d is the degree of the differencing factor, and B is the backshift (lag) operator.

Autoregressive models are used in time series analyses. These represent the current value of the process as a finite aggregrate of the previous values of the process and a shock a_t. Autobox uses polynomials of the form:

$$(1 - \Phi_1 B^1 - \Phi_2 B^2 - \Phi_3 B^3 - \ldots\ldots\ldots\ldots\ldots \Phi_p B^p) \tag{5}$$

where: Φ_1 to Φ_p are the parameter values of the polynomial,
B is the backshift operator for its autoregressive factors.

The values of the autoregressive factors (Φ_1 to Φ_p) need not all be non-zero. A zero parameter value indicates that the parameter is not included in the polynomial. Autobox will use as many factors as is appropriate.

Autobox will also use as many moving average factors as is appropriate. These represent the current observation as a function of a finite number of previous observations. Each moving average factor is a polynomial of the form:

$$(1 - \theta_1 B^1 - \Phi_2 B^2 - \Phi_3 B^3 - \ldots\ldots\ldots\ldots\ldots \Phi_q B^q) \tag{6}$$

where: θ_1 to Φ_q are the parameter values of the polynomial.

The values of the θ_p to Φ_q need not all be nonzero. A zero parameter value indicates that the parameter is not included in the polynomial.

6.5 Exception Analysis

The analysis of road data was done using an approach referred to in Autobox as **exception analysis**. In establishing the trend, two types of situations which give "unusual values" were considered:

- **Pulses** are unusual events in the data. For example, the measurements in one year may be incorrect due to improper calibration of the vehicle. These are commonly referred to as outliers but they could also be inliers[10].

[10] Consider the time series 1,9,1,9,17,9,1. Most techniques would clearly identify the value 17 as an outlier. However, the same series but with 1,9,1,9,5,9,1 has the value 5 as an inlier which can be defined as values that are "too normal or too close to the mean".

- **Level shifts** arise when there is a shift in the magnitude of the data, for example the roughness after shape correction will be much lower than before.

Autobox develops an autoregressive model which is used to identify the unusual values. This exception analysis is suited to the objectives of this project for assessing the quality of road data since its output can be used immediately in a quality control situation. However, the small sample sizes and scatter in the data present a challenge to the analytical capabilities.

6.6 Application of Autobox Exception Analysis to Sample Roughness Data

To evaluate the capabilities of the Autobox exception analysis routine, a sample data set of 20 x 100 m sections of roads were selected from the databases: 10 for State Highways and 10 from an RCA. The data were selected so that they covered the full range of data present: little variability, high variability, large pulses (inliers and outliers), and level shifts due to maintenance. Figure 6.1 shows the data used in the analysis. It is plotted in Figures 6.2 and 6.3.

NOTE: It must be emphasised that the data used in the analysis was not typical of State Highways or local authorities. It was selected specifically because it had features which would test the capabilities of the Autobox software.

HISTORY OF SERIES

Exit

	SERIES	1992	1993	1994	1995	1996	1997	1998	1999	2000
1	*12-17-A	90	90	78	74	71	76	84	83	85
2	*12-17-B	74	74	68	80	68	63	83	82	82
3	*12-17-C	69	65	69	89	64	96	124	133	139
4	*12-17-D	96	94	94	99	110	108	121	109	112
5	*12-17-E	101	112	126	108	118	117	109	116	107
6	*12-17-F	74	80	78	113	115	121	130	133	70
7	*12-17-G	64	76	94	74	79	82	83	79	72
8	*12-17-H	94	90	108	92	95	89	98	97	66
9	*12-17-I	98	99	103	134	123	136	153	133	72
10	*12-17-J	84	90	94	117	122	133	129	141	68
11	*BOUN-5K	65	60	53	59	55	52	55	58	61
12	*BOUN-5L	52	52	63	65	39	38	70	74	77
13	*BOUN-5M	43	44	56	54	50	57	73	77	80
14	*BOUN-5N	47	44	53	50	50	49	57	60	63
15	*BOUN-5O	47	44	60	61	60	54	53	56	58
16	*MOSS01T	52	44	49	42	40	43	51	54	56
17	*MOSS-1P	61	98	71	56	57	41	46	48	51
18	*MOSS-1Q	47	60	56	48	47	42	48	50	53
19	*MOSS-1R	47	52	56	45	40	42	49	51	54
20	*MOSS-1S	56	52	49	52	47	43	49	51	54

Figure 6.1 **Sample data for testing Autobox.**

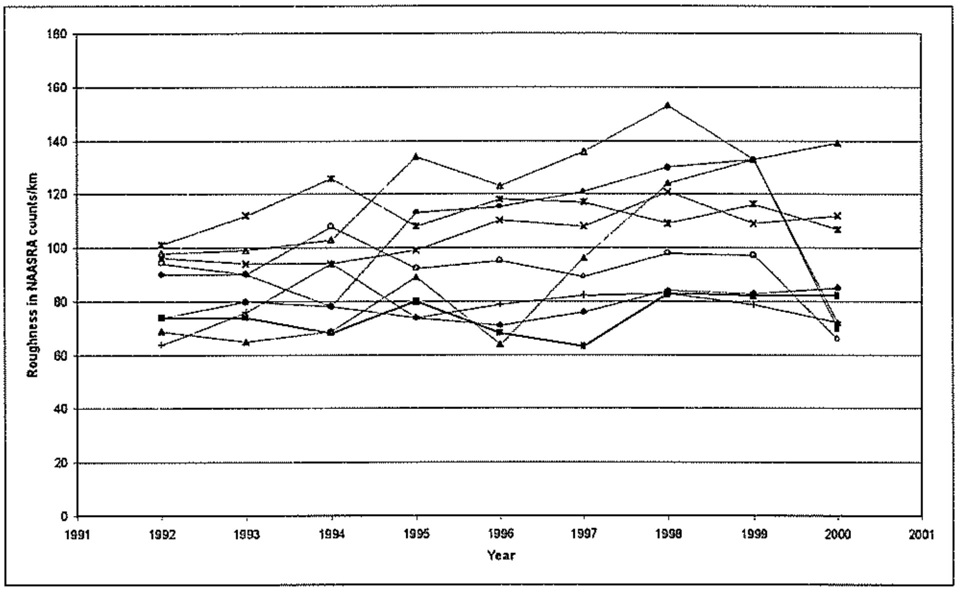


Figure 6.2 Sample data from State Highways.

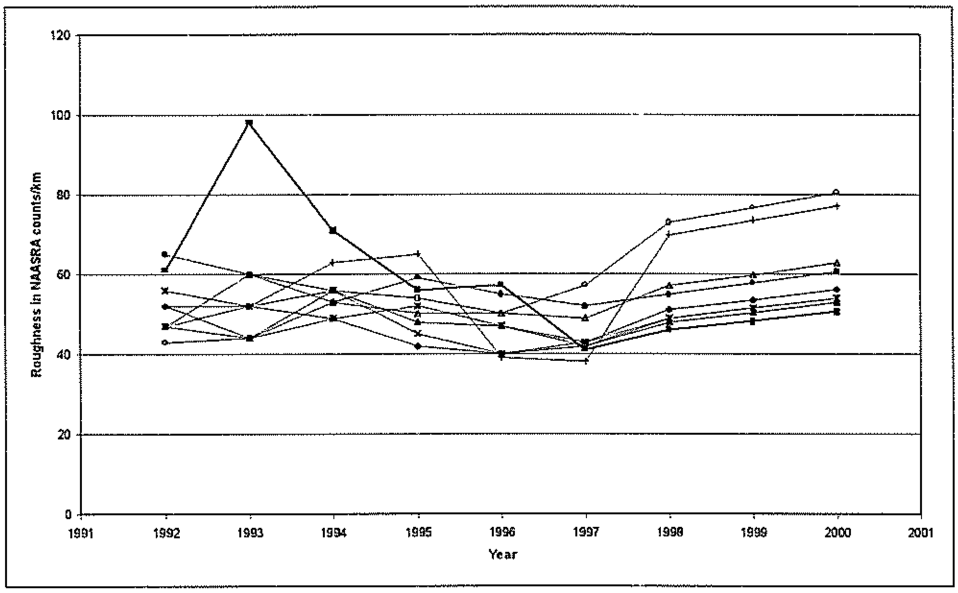


Figure 6.3 Sample data from an RCA.

The exception analysis has a series of reports. The most pertinent to the road data quality analysis are the **Pulse Alpha Values** and the **Level Shift Alpha Values**. These identify pulses or level shifts and how extreme they were in terms of a statistical anomaly. Values of 0.10 or less means that there is (1 - α) probability that the value is significant.

Figures 6.4 and 6.5 show the results of the pulse and level shift testing for the sample data. Table 6.1 shows the pulses and level shifts that were >90% probable against the original data.

PULSE ALPHA VALUES

Exit Sort

	SERIES	NEG	POS	TOT	R2	1992	1993	1994	1995	1996	1997	1998	1999	2000
1	*12-17-A	0	1	1	0.23		0.14312							
2	*12-17-B	1	0	1	0.36						0.05545			
3	*12-17-C	1	0	1	0.89					0.00415				
4	*12-17-D	0	0	0	0.82									
5	*12-17-E	0	1	1	0.45			0.02898						
6	*12-17-F	0	0	0	0.32									
7	*12-17-G	0	1	1	0.51			0.01745						
8	*12-17-H	1	0	1	0.75									0.00121
9	*12-17-I	1	0	1	0.43									0.03288
10	*12-17-J	1	0	1	0.91									0.00011
11	*BOUN-5K	0	1	1	0.44	0.03134								
12	*BOUN-5L	1	0	1	0.29						0.09309			
13	*BOUN-5M	0	1	1	0.82			0.10474						
14	*BOUN-5N	0	0	0	0.5									
15	*BOUN-5O	1	0	1	0.45		0.02827							
16	*MOSS01T	1	0	1	0.26					0.11674				
17	*MOSS-1P	0	1	1	0.73		0.00161							
18	*MOSS-1Q	0	1	1	0.46		0.02593							
19	*MOSS-1R	1	0	1	0.34					0.06709				
20	*MOSS-1S	1	0	1	0.5						0.01925			

Figure 6.4 Significance of Pulse testing.

LEVEL SHIFT ALPHA VALUES

Exit Sort

	SERIES	NEG	POS	TOT	R2	1992	1993	1994	1995	1996	1997	1998	1999	2000
1	*12-17-A	0	0	0	0.23									
2	*12-17-B	0	0	0	0.36									
3	*12-17-C	0	0	0	0.89									
4	*12-17-D	0	1	1	0.82					0.00035				
5	*12-17-E	0	0	0	0.45									
6	*12-17-F	0	1	1	0.32					0.07831				
7	*12-17-G	0	0	0	0.51									
8	*12-17-H	0	0	0	0.75									
9	*12-17-I	0	0	0	0.43									
10	*12-17-J	0	0	0	0.91									
11	*BOUN-5K	0	0	0	0.44									
12	*BOUN-5L	0	0	0	0.29									
13	*BOUN-5M	0	0	0	0.82									
14	*BOUN-5N	0	1	1	0.5						0.01912			
15	*BOUN-5O	0	0	0	0.45									
16	*MOSS01T	0	0	0	0.26									
17	*MOSS-1P	0	0	0	0.73									
18	*MOSS-1Q	0	0	0	0.46									
19	*MOSS-1R	0	0	0	0.34									
20	*MOSS-1S	0	0	0	0.5									

Figure 6.5 Significance of Level Shift testing.

Appendix E shows the predicted time series trends from Autobox for each section. Of the 20 x 100 m sections tested, only three were found to be suitable for modelling as a time series, two from the State Highways and one from the Local Authority. The State Highway results are shown on Figure 6.6.

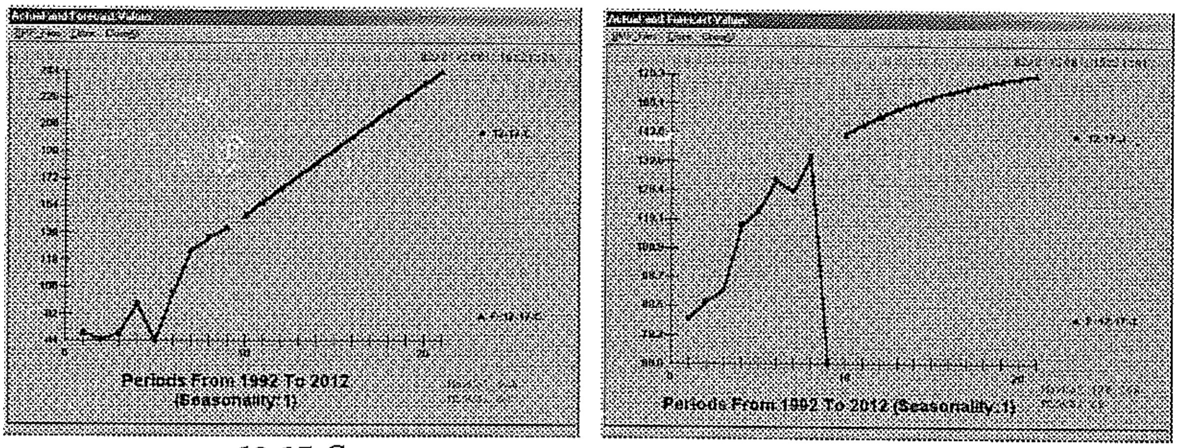
In both 12-17-C and 12-17-J the software successfully identified pulses (observations 5 and 9 respectively on Figure 6.6). These were ignored in establishing the trend line. The results for 12-17-J are particularly pertinent for the objectives of this project because, if this data had been provided by a contractor, it would have been identified as unusual.

Table 6.1 Pulse and Level Shifts identified in Autobox analysis.

Year	12-17-A	12-17-B	12-17-C	12-17-D	12-17-E	12-17-F	12-17-G	12-17-H	12-17-I	12-17-J	MOSS01 T	MOSS-1P	MOSS-1Q	MOSS-1R	MOSS-1S	BOUN-5K	BOUN-5L	BOUN-5M	BOUN-5N	BOUN-5O
1992	90	74	69	96	101	74	64	94	98	84	52	61	47	47	56	65	52	43	47	47
1993	90	74	65	94	112	80	76	90	99	90	44	98	60	52	52	60	52	44	44	44
1994	78	68	69	94	126	78	94	108	103	94	49	71	56	56	49	53	63	56	53	60
1995	74	80	89	99	108	113	74	92	134	117	42	56	48	45	52	59	65	54	50	61
1996	71	68	64	110	118	115	79	95	123	122	40	57	47	40	47	55	39	50	50	60
1997	76	63	96	108	117	121	82	89	136	133	43	41	42	42	43	52	38	57	49	54
1998	84	83	124	121	109	130	83	98	153	129	51	46	48	49	49	55	70	73	57	53
1999	83	82	133	109	116	133	79	97	133	141	54	48	50	51	51	58	74	77	60	56
2000	85	82	139	112	107	70	72	66	72	68	56	51	53	54	54	61	77	80	63	58

> 90% Probability Pulse
> 95% Probability Pulse
> 99% Pulse Pulse

Shift



12-17-C *12-17-J*

Figure 6.6 Successful Time Series analyses of State Highway results.

While Autobox was unable to fit data to most of the time series, as shown in Appendix E, these all had multiple readings which were "unusual", and this precluded fitting a sensible trend. Of more importance to this project was Autobox's ability to identify anomalies in the data, specifically pulses and level shifts.

Reviewing the data from Table 6.1, it can be seen that, considering the data, the software did a fair job of identifying pulses and level shifts in the data. For example, on Sections 12-17-H, 12-17-I and 12-17-J it correctly identified the maintenance activity in the year 2000 as a pulse. However, it missed the maintenance on 12-17-F in the same year, which was surprising given the magnitude of the roughness reduction. This could have been caused by it identifying a shift in the data in 1996.

Autobox clearly identified other pulses in a number of instances, for example 12-17-C (1996); 12-17-E (1994); 12-17-G (1994), MOSS-1P (1993). Some of the other pulses identified (e.g. 12-17-B in 1997) are confusing and, without resorting to a detailed review of the statistical analysis, it is not possible to ascertain why these were identified as such.

The identification of level shifts was less successful, but this is probably because of the small sample sizes being dealt with, and the amount of noise in the data. For example, 12-17-C had a level shift in 1998 but this could have been clouded by the low pulse in 1996.

The analysis with Autobox showed that it is a powerful tool which can overcome many of the problems that exist with road data, for example small sample sizes and high noise. The ability of the software to identify pulses is quite important since that is the goal of the data quality project: to be able to identify anomalous results when data are provided by a contractor. However, for this to work an improvement in the quality of data is needed, and this will represent the greatest challenge.

7. Conclusions

As described in Chapter 1 of this report, data are considered to be the required quality:

> *...if it satisfies the requirements stated in a particular specification and the specification reflects the implied needs of the user.*

The review of the data conducted here suggests that the data may not always be of the quality required to make correct decisions with regard to managing roads. This lack of quality may be related to many different factors, but appear to come down to two principal factors:

- **Precision**. Under controlled conditions roughness meters were found to have standard deviations of 2-3%. It is not known what the precision would be under field surveys, but given that the vehicles in the field operate for long hours, under prevailing climatic and traffic conditions, one would anticipate that the precision would be somewhat lower. This may be one of the reasons behind some of the "noise" in the time series.
- **Quality Assurance**. There were notable instances where the data showed the effects of a lack of quality assurance. These ranged from improper location referencing, to significant variations in the roughness between years which could not be prescribed to a lack of precision.

The State Highway roughness data showed much more consistency and less variability than the local authority data. This is probably a reflection of the high level of quality assurance imposed on the State Highway contractors.

The current practice of looking at average conditions along sections of road, or for entire areas, for quality assurance purposes is inadequate since it masks what can be significant variations within a section because of measurement or operational errors. For example, one local authority had a decrease in its average roughness between two years when looking at all roads, yet on a number of individual sections there were increases in the roughness which were of such a magnitude that they should have been rejected as invalid.

The RAMM visual rating data showed a great deal of variation between years, but this could be due in part to the impact of minor maintenance activities in addition to problems with the data. Where trends were observable, often a lot of scatter was recorded which suggested a lack of precision in the measurements, not unexpected given that they were done through visual observation.

This failure to record maintenance activities in RAMM, in a manner appropriate for time series analysis, is the largest single impediment to instituting a robust system for assessing data quality. It leads to discontinuities in the time trends which are either incorrectly rejected as outliers or which cause problems in the analysis.

The analysis of the roughness data using Autobox showed that it is possible to achieve the original objective of the project, identifying inconsistent data, but that more work needs to be done to find a solution appropriately tailored to analysis road data. The combination of noise in the data, small sample sizes, and maintenance activities create a complicated framework for any type of analysis to correctly identify inconsistent data. It is likely that, by including a spatial element to the analysis, one would have additional information which could be used to overcome some, if not many, of these problems.

Future analyses of data will be facilitated by the time series databases established in this project. This was a mammoth and complicated undertaking but a procedure has been established which can be used in future to create databases for similar analyses. The databases are stored on two CDs and are available to other researchers on request.

8. Recommendations

The principal recommendations from this research are as follows:

- The way in which maintenance activities are recorded in RAMM should be reviewed, with the objective of establishing a procedure which will facilitate time series analysis of the data. Other changes to the architecture of RAMM should be considered to assist with exporting the data for quality assurance evaluations.
- The precision of the various road survey measurements that are in use needs to be established. In discussing the application of data with practitioners, an expected precision is associated with the data, which may not in fact be appropriate. Knowing the precision of measurements under typical survey conditions would also facilitate the quality assurance process, insofar as it would indicate the expected change from year-to-year due to measurement variability
- While the time series analyses with Autobox showed that it is possible to identify unusual readings, the small sample sizes and noise in the data helped to reduce the reliability of the analysis. For practical use, too many data items would be identified as questionable using this method. It would be better to use a different approach, for example one which took into account the spatial element of the data, than a Box-Jenkins time series approach.

  Other outlier techniques are available and these should be explored. The databases developed in this project would provide the ideal springboard for developing this alternative statistical approach. The work should be developed into a stand-alone software application which could be used by all RCAs for assessing their data quality.

9. References

Abate, M.L., Diegert, K.V., Allen, H.W. 1998. A hierarchical approach to improving data quality. *Data Quality Journal 4 (1).*

BECA. 1996. *Trial of high speed data collection vehicle – validation of outputs.* Report to Transit NZ, by Beca Carter Hollings and Ferner Ltd., Auckland.

Bennett, C.R. 1998. Evaluation of a transverse profile logger. *Proceedings 4th International Conference on Managing Pavements*, Durban.

Box, G.P., Jenkins, G.M., Reinsel, G.C. 1994. *Time series analysis: forecasting and control.* Prentice Hall, London.

Karamihas, S.M., Gillespie, T.D., Perera, R.W., Kohn, S.D. 1999. Guidelines for longitudinal pavement profile measurement. *NHCRP Report 434*, Transportation Research Board, Washington, DC.

McGhee, K. 2000. Quality assurance of road roughness measurements. *Report VCTR 00-R20RB*, Virginia Department of Transport, Virginia, USA..
http://www.vdot.state.va.us/vtrc/briefs/00-r20rb/Road_Roughness.htm

Millard, R. 1993. *Road building in the tropics.* Transport Research Laboratory (TRL), Crowthorne, UK.

NZIHT. 2000. *RAMM condition rating – sealed roads.* New Zealand Institute of Highway Technology, New Plymouth.

Paterson, W.D.O. 1987. *Road deterioration and maintenance effects: models for planning and management.* World Bank Publications, Washington, DC.

Paterson, W.D.O., Scullion, T. 1990. Information systems for road management: draft guidelines on system design and data issues. *World Bank Technical Paper INU 77*, Infrastructure and Urban Development Department, The World Bank, Washington, DC.

Transfund New Zealand (Transfund). 1997. *RAMM Road Condition Rating and Roughness Manual.* Transfund New Zealand, Wellington.

Appendices

Appendix A. Precision of Roughness Measurements

In order to assess the precision of roughness measurements, data were collected from measurements in New Zealand using a response-type roughness meter, and from the published literature for profilometers.

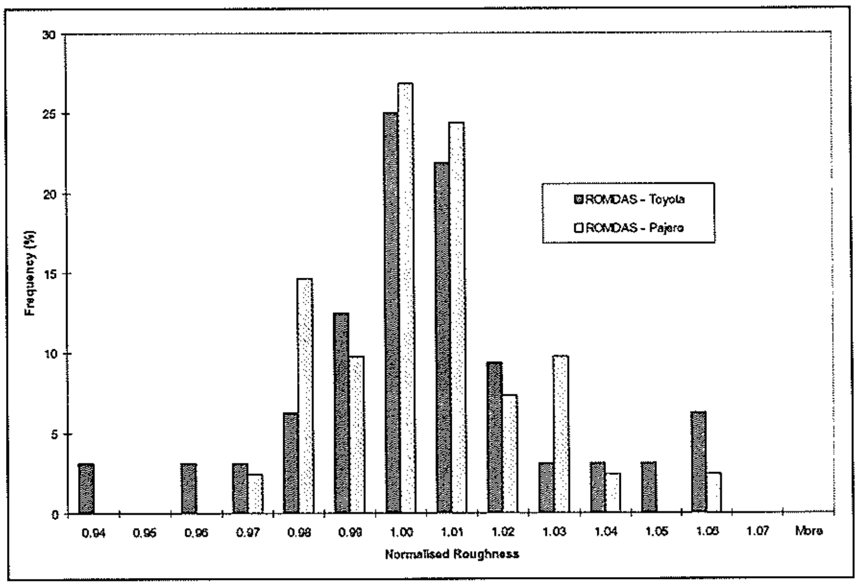
The data for the repeatability analysis came from calibrations using the two HTC ROMDAS survey vehicles: a Toyota van and a Mistubishi Pajero, hereinafter referred to as the Toyota and Pajero.

The data were collected in repeat runs over a series of test sections in August 2000 (Pajero) and November 2000 (Toyota). Different sections were used in each study, along with different drivers. Measurements were made at 50 km/h and 80 km/h.

The data consisted of the total raw roughness for both wheelpaths at each site, with a minimum of 2 runs being made at each site. A total of 41 site runs were available for the Pajero; 32 for the Toyota.

As suggested by Karamihas et al. (1999), the normalised raw roughness was adopted as the measure of repeatability. This was calculated by dividing the roughness from each run by the mean of all runs for that site. Thus, if an individual test run produced a roughness of 3000 counts/km and the average for all runs was 3050 counts/km, the normalised roughness was 0.984. The advantage of using this measure is that it is independent of the magnitude of the roughness and can therefore be used across all sites.

Figure A1 Histogram showing the distribution of the normalised roughness data.



The repeatability of the ROMDAS measurements had data in the following ranges:

Pajero – 0.965 – 1.052
Toyota – 0.935 – 1.058

Karamihas et al. (1999) used three criteria for assessing the repeatability of profilometers: standard deviation, and the percentage of measurements within 2% and 5%. Table A1 shows the ROMDAS data using these criteria.

Table A1 ROMDAS data using criteria of Standard Deviation and percentage of measurements within 2% and 5%.

Vehicle	Std Dev.	% within 2%	% within 5%
ROMDAS - Pajero	1.85	68.3	97.6
ROMDAS - Toyota	2.57	68.8	90.6

For profilometers, repeatability results were obtained from Karamihas et al. (1999), which was based on data from a 1993 RPUG experiment with 33 profilometers; and McGhee (2000) who used a modern 32 kHz laser profilometer. The results are shown in Table A2.

Table A2 Repeatability results for different profilometers.

Vehicle	Std Dev.	% within 2%	% within 5%
Karamihas et al. (1999) [average of 33]	4.49	49.7	81.5
McGhee (2000)	1.89	76.7	97.5

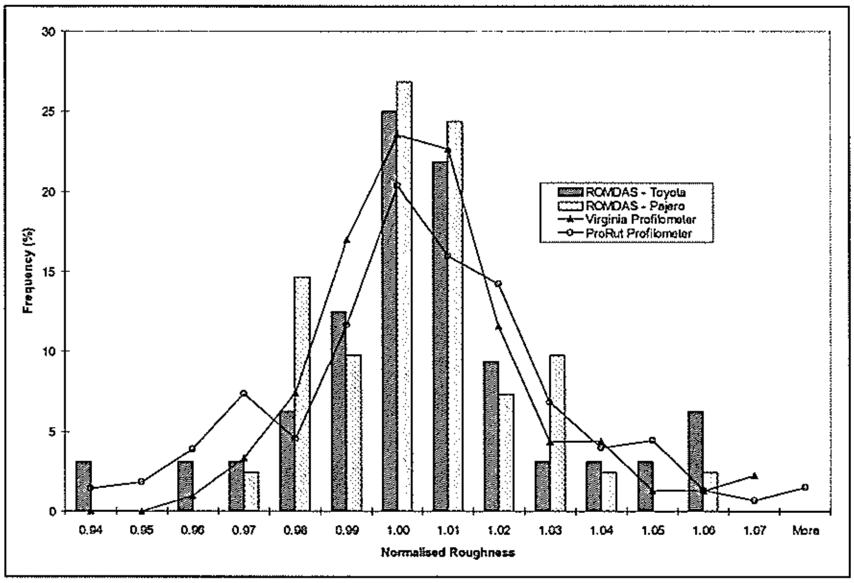
Of the 33 profilometers in the Karamihas et al. (1999) study, it was clear that those using ultrasonic principles had problems with repeatability. The laser and optical profilometers performed much better. This is confirmed in the following Table A3 which shows the results against the types of instruments from the RPUG study. The ROMDAS response-type meter is also included.

Table A3 Results from RPUG study of Karamihas et al. (1999).

Instrument	Std Dev. (%)	% within 2%	% within 5%
Optical	2.95	72.6	95.0
ROMDAS	2.17	68.5	94.5
Laser	3.23	58.3	91.2
Ultrasonic, Commercial	5.32	42.2	77.7
Ultrasonic, Agency-built	6.47	36.4	67.1

The histogram in Figure A2 shows the ROMDAS distribution along with distributions from a ProRut profilometer (Karamihas et al. 1999) and the Virginia profilometer (McGhee 2000). With the exception of a few outliers with the Toyota, the response-type systems show equivalent repeatability to the profilometers.

Figure A2 Distributions for ROMDAS, and two profilometers.



To further investigate the precision of the ROMDAS Bump Integrator (BI), data were obtained from calibrations undertaken by ROMDAS users in different countries. For each speed and site the raw roughness was divided by the mean roughness to obtain the normalised roughness. These normalised roughnesses were then analysed to investigate the percentages within 2% and 5%. Each country represents a unique driver-vehicle combination. The results are given in Table A4.

Table A4 Results of calibrations from different countries.

Vehicle	Std Dev.	% within 2%	% within 5%
Tonga (n=48)	2.7%	56.3	93.8
Indonesia (n=144)	3.4%	56.3	90.9
Malaysia (n=50)	3.0%	64.0	88.0
Laos (n=72)	5.2%	27.8	65.3

These data show that wide variations can be expected when using the same instruments, and that these differences are a function of the driver and vehicle combination.

On the basis of the analysis here, the assumption is that response-type measurements have a standard deviation of 3% and laser profilometers of 2%.

Appendix B. Preparing Data for Analysis

The source data from the RAMM database was processed for analysis. This was done using the HTC Information Management System (HIMS) and Microsoft Access. The steps consisted of:

- Import source data,
- Organise source data into different year/lane,
- Transform treatment data (surface and pavement) into other data of different year/lane,
- Transform data into 100 m new sectioning format,
- Organise data into final format for analysis.

Each of these steps is described in detail below. Depending on the data available, the actual steps may differ. The example uses the Northland PSMC source database.

The first step is to import source data into a HIMS source database. Source data from RAMM is preferred in the pipe-delimited (|) format and a data import schema must be created for import. Once imported, this HIMS source data is organised into another HIMS source database with data in the required final format of different year and lane.

Other source data format including tab-delimited data with various heading names and different number of fields, often needs to manually match each field for importing, and then the data organised from imported tables, although direct organisation of data into year/lane is possible as well. The following description applies to pipe-delimited source data format.

❑ Create HIMS database and HIMS source database

Start HIMS, and create a new HIMS database named HIMSPSMC.mdb.
Add a new source database called PSMC Data.
Figure B1 is the screen for defining a source database in HIMS.

❑ Define table structure for source database

To define the table structure in source database, select Define Data Structure, and then select Display Existing Tables. The screen in Figure B2 will appear.

Because the pipe-delimited data source does not have column headers, it is necessary to define structures for each source table. The following is the source table name list. The format of the display is: table description (file name, table name)→table name in HIMS source database).

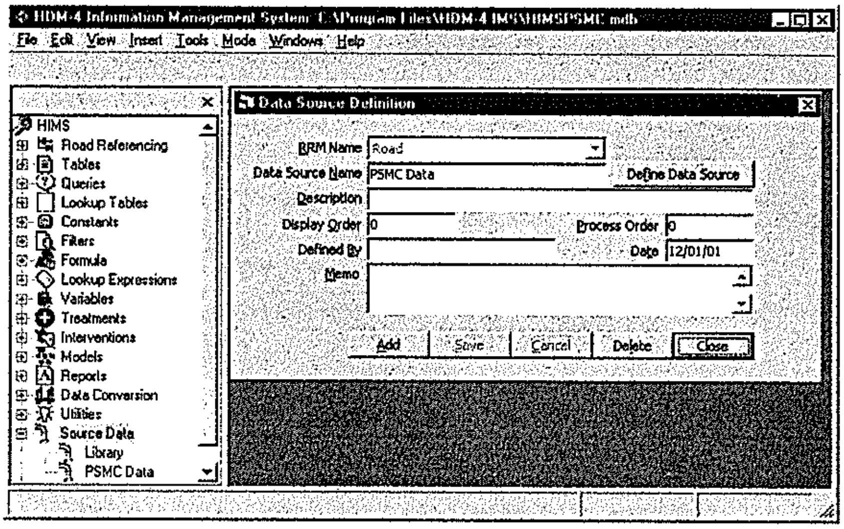


Figure B1 Defining database for source data.

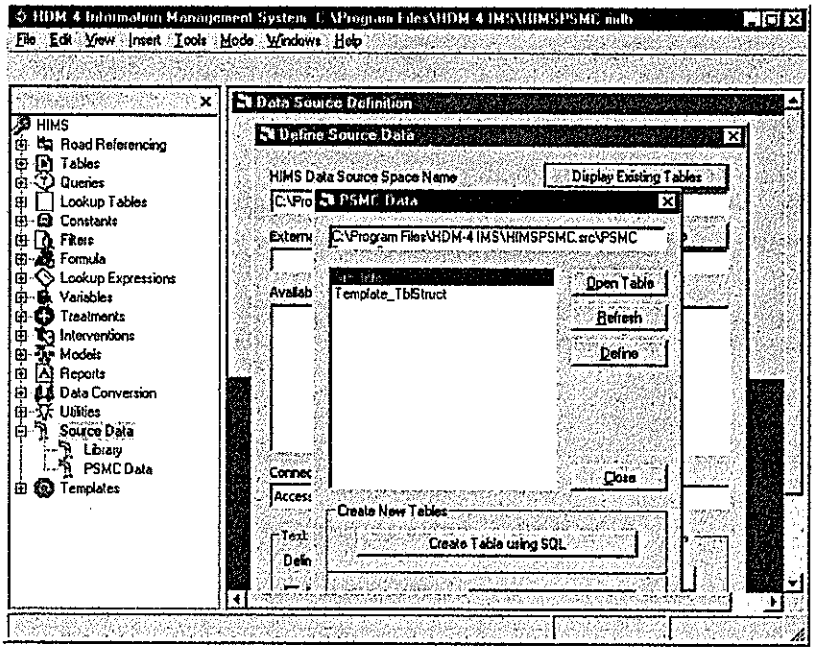


Figure B2 Display existing tables in source database.

Carriage Way (Carr_way.unl, carr_way table) → carr_way#asc
Road Name (Roadname, RoadNames table) → roadbanes#asc
Geometry (HSgeometry file, HSD_Geometry table) → geometry#asc
HSD-roughness (HSDRough.unl, HSD_Rough file) → HSD-Roughness#asc
HSD-rutting (HSDrutt.unl file, HSD_Rutting table) → HSD-Rutting#asc
HSD-Skid (Skidres.unl file, Skid_Resistance table) → HSD-Skid#asc
HSD-texture (HSDtext.unl file, HSD_Texture table) → HSD-Texture#asc
Rating (Rating.unl file, Rating table) → Rating#asc
Roughness (Rough.unl file, Rough table) → Roughness#asc
Top Surface (TopSurf.unl file, Top-Surface table) → Surface#asc
Pavement Structure (p_struct.un file, Pave_Structure table) → Pave_Structure#asc

In the case of RAMM export, data will be in different format and have different file names, numbers of fields and field names. Some data can be a combination of several tables, e.g. Pavement table.[11]

Other information can be included, but is not currently used. They are:

- Traffic data
- Treatment Length data

To define the table structure the information in the road.sql file from RAMM Informix unloading is used. Open it in WordPad, and find the appropriate table name, then copy and paste the definition and create the table in HIMS Source database. The following is an example.

1. Open road.sql in WordPad. This is the data definition file for the RAMM database which includes all the table definitions.
2. Use the search function to find the table definition. The table name is given above (e.g. carr_way, or rating). The definition part for a table is from "Create.. and then "(" and up to the end ");" (see Figure B3).
3. Copy the text to the Windows clipboard.
4. In the HIMS management panel, select "Display Existing Tables", and then select "Creating Table Using SQL Definition".
5. Paste the clipboard text to the text box.
6. Click Convert String to convert the definition from Informix into Access SQL format.
7. The Informix style contains owner information such as "n9dar0". Delete this.
8. Change the table name change the table name to Rating#asc from Rating. In the first line is: "Create table 'n9dar0'.rating ( )", the actual translation in SQL DDL is: create a table named rating (after the ".") according to the definition inside the "( )", this table is owned by "n9dar0".
9. Click create table.

[11] RAMM has two versions: Unix and Windows. When data are unloaded from RAMM in Unix it includes all the sql and data files, which can then be loaded to a new database for transferring the files. All these files are .unl files and in pipe delimited format. The DDL file (data definition language for SQL) road.sql is also included. Enough information is provided to build a complete database.

When exporting from RAMM for Windows alias names are used for each field. This creates a number of problems. Firstly, the names are different between versions. There will also be a different number of fields included in the table, and sometimes fields may appear twice in a table. Furthermore, the table can be called any name depending on the what the user decides. As a consequence of this, importing data from RAMM for Windows for manipulation is very problematic.

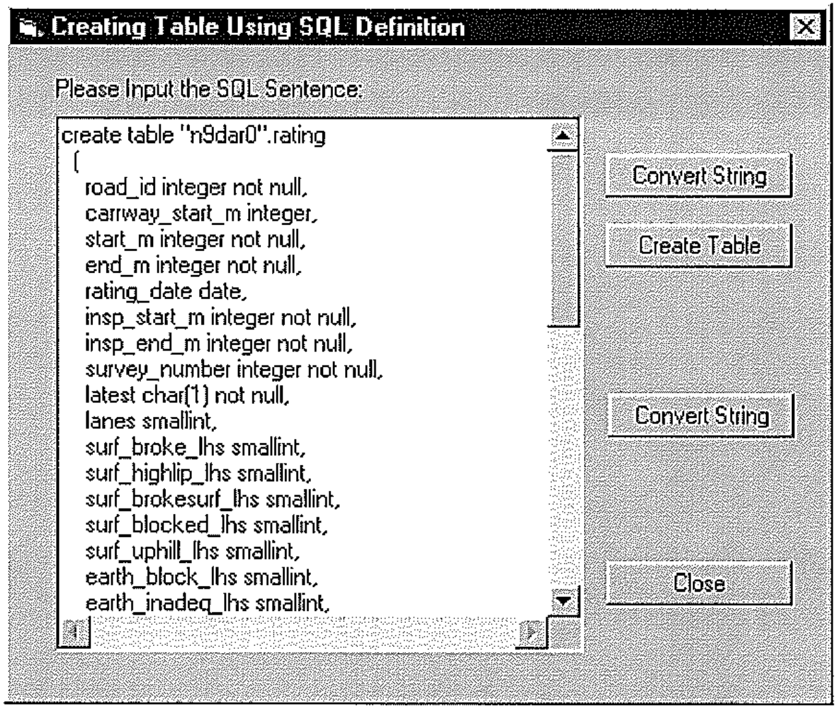


Figure B.3 Creating Table Using SQL Definition.

Once all the table structures are defined, close this window.

This source database can be saved as a template for later use with other data set. This is the recommended process since this step can then be skipped and the data can be directly imported into a source database.

- **Import source data into source database**

The source data can now be imported into the corresponding table in the source database. To do this, in the **Define Source Data** screen, select the place where all the source data files are kept by selecting any file (all source data files will be displayed in the left list box) and then select those tables you want to import. Be sure to give the right target table name.

To import PSMC source data, Select the **Define Data Source** screen (or select Define Structure for PSMC Data from the Edit menu, or drop down menu, see Figure B.4). Then select Browse to select the source data files, all files will appear in the list boxes. Select the tables to import from the Available Source Data Tables, and Edit the target table name in Edit Saved Table Name box, select Delimited Character as PipeDelimited, and then select OK to import data. This is shown in Figure B 5.

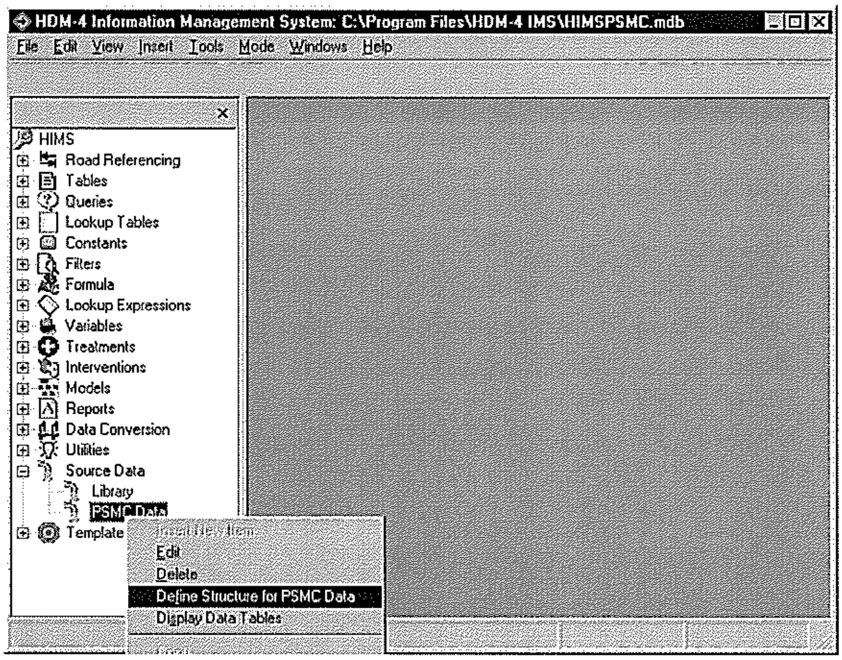


Figure B.4 Activate Define Source Data Screen.

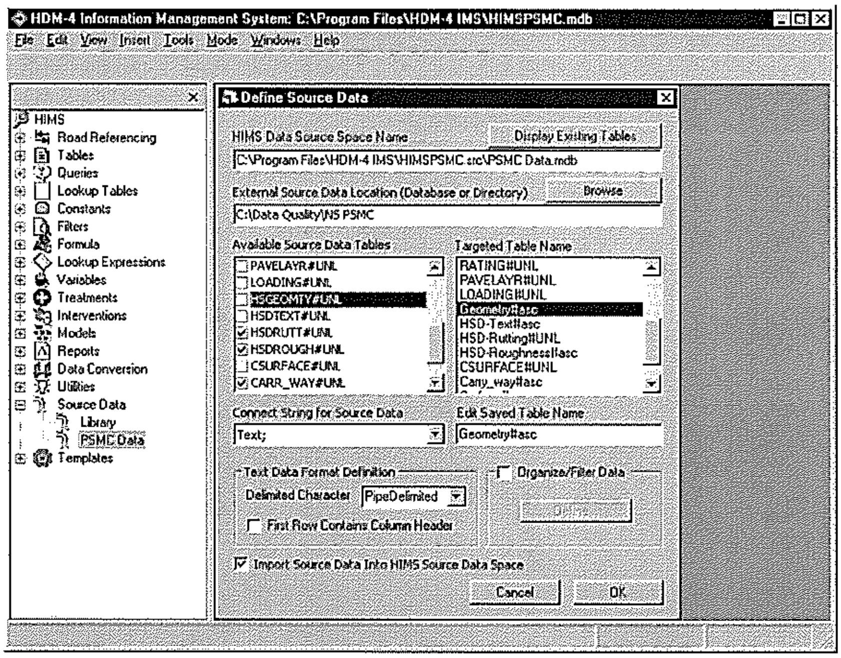


Figure B 5 Selecting Source Data Files (Tables) to Import.

During the import process you will be prompted to create or overwrite an existing schema file if one already exists. Be sure to select **No-Overwrite** (i.e. No in the following screen), and select No in order not to delete the existing tables. This means that all the selected tables will be imported to the target tables.

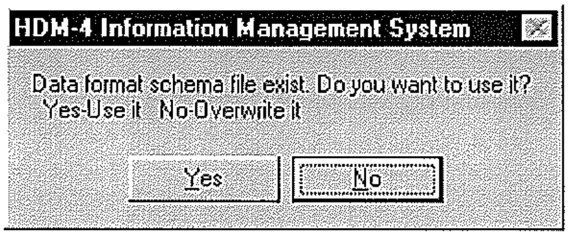


Figure B 6 Select No to Create New Schema for import.

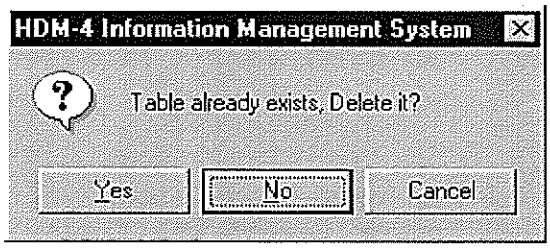


Figure B.7 Select No to Append Data.

After importing, all imported source files will disappear from the list boxes.

In the RAMM database, data from different years and different lanes are kept together for different road segments. Importantly, the start and end chainages are not fixed for data of different years/lanes. This makes the data very difficult to use in time series analyses such as are required for data validation.

In order to make the data into a comparable and consistent format, one needs to organise the data initially into different year-lanes, and then transform the data into the same sections. This is done as follows:

For roughness, the data for each year-lane will have the following format:

Roughness#asc_1992_L1
Roughness#asc_1992_L2
...
Roughness#asc_1992_R1
Roughness#asc_1992_R2
...

For Rating, data will apply to all lanes, so it is not grouped into different lane, i.e.:

Rating#asc_1992
Rating#asc_1993
...

For treatment data, there was not enough information to separate the data into individual lanes (i.e. L1 or L2), so we only separate them into the left and right lane for each year (i.e. L and R). This is adequate since analyses were only carried out for data in lane 1 (left lane 1, and right lane 1). The format for treatment data is:

Surface#asc_L_1992
Surface#asc_R_1992
...
Pavement#asc_L_1992
Pavement#asc_R_1992
...

The number of years and lanes depends on the actual data available. Consequently, there are differences in the number of year/lane tables in the databases.

The following steps describe how the data were organised:

- **Transform Carriageway width into treatment data tables** (surface table and pavement structure table)

  Select **Tools|Utilities|Transformation** (or in HIMS management panel, expand the Utilities entry, run the Transformation utility). Define the source data properties and the field to be transformed (see Figure B 8).

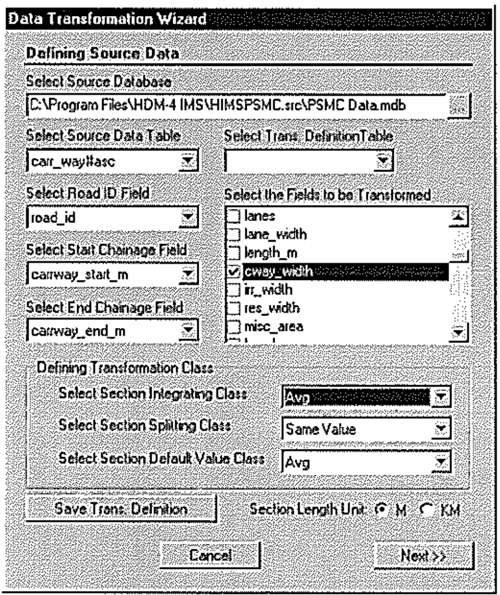


Figure B.8 Define Data Properties for Source Data

Select **Next** and then define the Destination data properties (see Figure B 9). Then select **Finish** to do the data transformation.

Transform the carriageway width into surface table in the same way.

Please note, either Access or the HIMS could be used to make a query to do this. HIMS can only apply queries on an HIMS database instead of the source database.

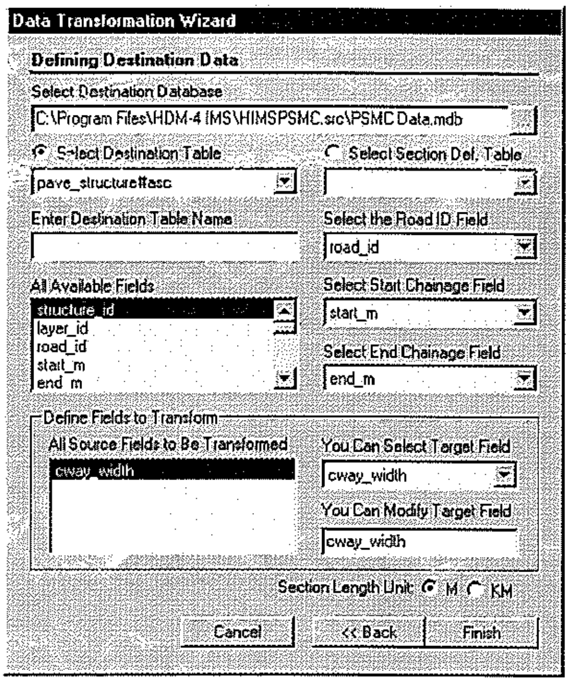


Figure B9 Define Data Properties for Destination Data.

- **Verify/modify lane data: in the data source**, the name for the lane may be different but for analytical purposes needed to be made consistent. For example, in some records, it may be recorded as L1, or L 1, or Left Lane, or Left Lane 1 etc. Creating a consistent name is done as follows:

  - Select PSMC Data in the source data entry in HIMS management panel, right click on it and select Open Existing Tables to display all existing tables. Select the data tables from the list box and then click Open Table to view the table content. Then sort the field Lane (or direction in Roughness table), and check the value. If anything inconsistent is found, use the Assign function to modify the data (see Figure B10). For detailed operation instructions, please see the HIMS user manual. After this process one should only have lane value as L1, L2..., or R1, R2.., or B (both direction in case of only two lanes).

  - Create a new source database PSMC Data1 in HIMS as was done above when creating PSMC Data.

  - Define the Data Source for PSMC Data1 and import/organise the data form PSMC Data.mdb (created in the step Import Source Data). Open the screen, select the data source by locating the PSMC Data.mdb (under the HIMSPSMC.src subdirectory, in the same directory as HIMSPSMC.mdb HIMS database) All tables will be displayed in the list box (see Figure B11).

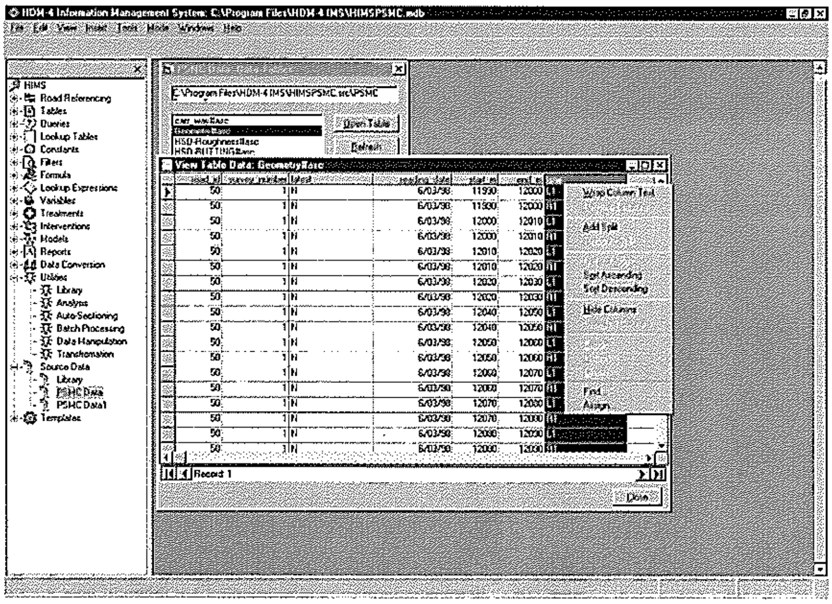


Figure B 10 Verify and Modify Source Data.

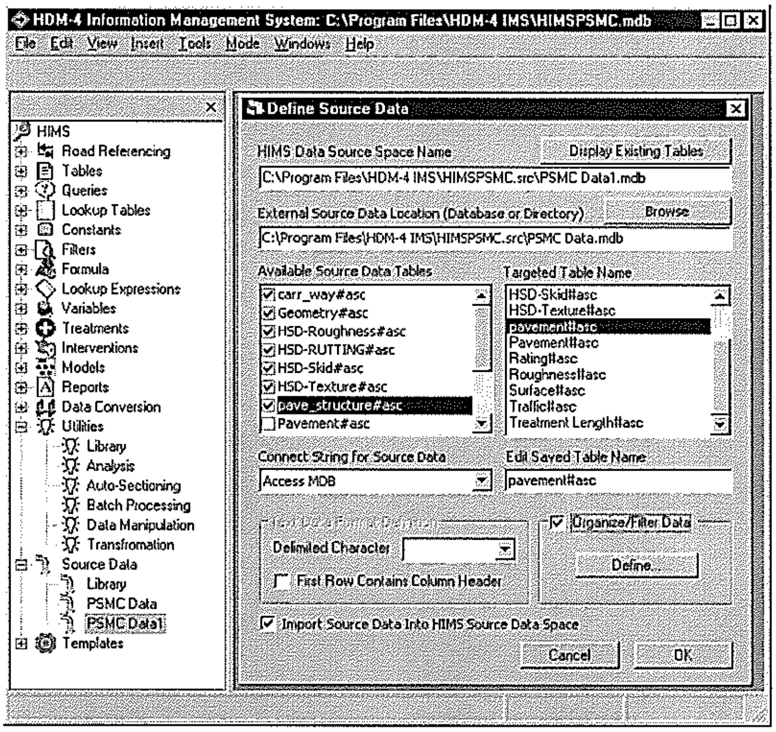


Figure B 11 Define Source Data for PSMC Data1.

Import most of the source data tables, and change the target table name for Pave_Structure#asc to Pavement#asc. Then check the box Organise/Filter data, and select Define to define the import properties. All selected tables will be displayed in the list box (see Figure B 12).

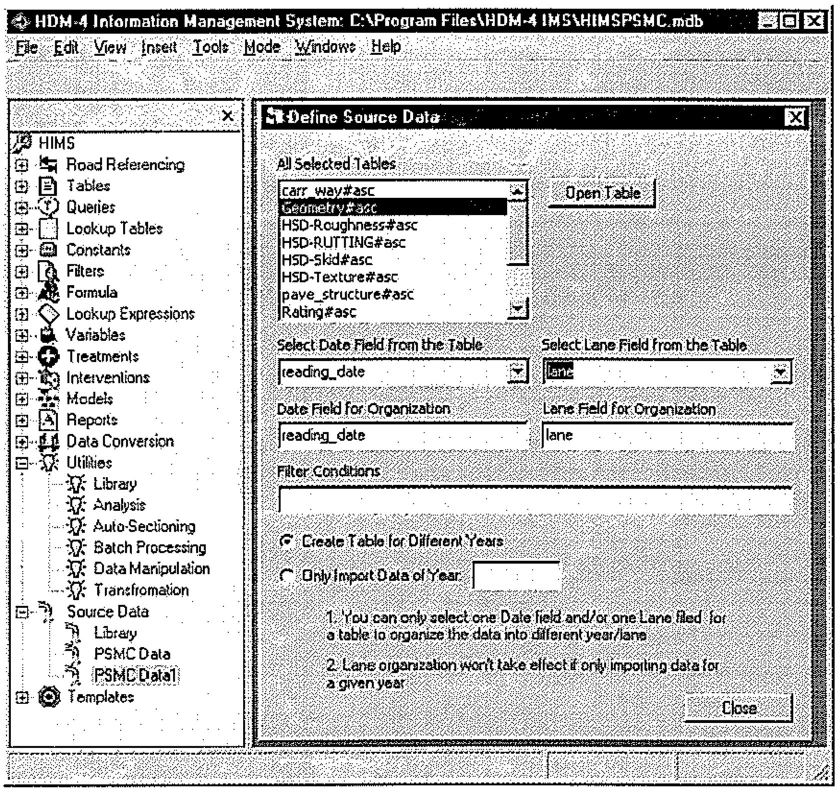


Figure B12 Define Import Properties for Tables.

For the carriageway table, there is no need to define anything. For all other tables, define the Date and/or lane to organise the data, and for treatment data, to define a filter to get the lane. This is described as follows:

- HSD survey data: reading_date, Lane
- Roughness data: reading_date, direction
- Rating data: only rating_date
- For carriage way data, traffic data, treatment length data: do not define
- For surface data: Surface_Date, and a filter to organise data into different left and right lanes. The filter is:

for left lane:
[Surface_Date]>#1/1/1990# and [Surf_Width]>2 and [Cway_Width]/2-[Surf_Offset]>2

for right lane:
[Surface_Date]>#1/1/1990# and [Surf_Width]>2 and [Surf_Offset]+[Surf_Width]-[Cway_Width]/2>2

Here only treatments after 1990 and with a treatment width > 2 m will be solved during transformation. Since there is no information on the number of lanes a road has, whether it is unidirectional or not, we treat all roads as two lanes and having two directions. If a road turns out to be unidirectional, then only half of the treatment data will be used; if a road has more than one left (or right) lane, this may be slightly skewed. The final verification of the analysis will correct this to some degree.

For the pave_structure data, same rule applies. The filter condition is:

For left lane:
[Layer_date]>#1/1/1990# and [layer_No]=1 and Width>2 and [Cway_Width]/2 – [Offset[>2

for right lane:
[Layer_date]>#1/1/1990# and [Layer_No]=1 and Width>2 and Width + offset –[Cway_Width]/2 >2

Here, only treatments in Layer 1 are considered.

For the surface and pavement table, two passes of the analysis are required to create both directions. For the left lane, it is done with other data sources, and then the data are re-imported for the right lane. This is a special situation, and we must change the External Source Data Location and then change back to get the tables.[12]

Once all the import properties have been defined, select Close to close the screen, and then select OK to import and organise all data. Figure B13 shows all the data organised in PSMC Data1.

- **Check and manually correct the data table** for data pertaining to both directions. After organisation, tables like Roughness_1992_B may appear. In this case, open the database in Access, and append the data into the L1 and R1 data table of the same year (e.g. append data from Roughness_1992_B into Roughness_1992_L1 and Roughness_1992_R1 etc.).

In order to compare data, create a new section for each road, and then transform all data against this sectioning. Select 100 m fixed length sectioning (the Roughness data is in 100 m interval, while the HSD data is in 10 m interval[13]). The following steps are for creating new road sections and transforming data:

---

12 This is required since the directory containing the source files is taken as a database in the HIMS, and all the data files as tables in the database. The program does a comparison: if you select the same file name (directory name) as the open database (or directory), it then uses the existing one instead of closing it and open it again. If the file name is different, then the old one will be closed, and a new database will be opened, and all its objects will be looped out.

13 It was originally planned to analyse the HSD data in 10 m intervals, however, it was found that there was significant scatter in the data which appeared due to the short sampling interval. On advice from Transit NZ it was aggregated to 100 m which also served to make it consistent with the NAASRA data collected by other RCAs.

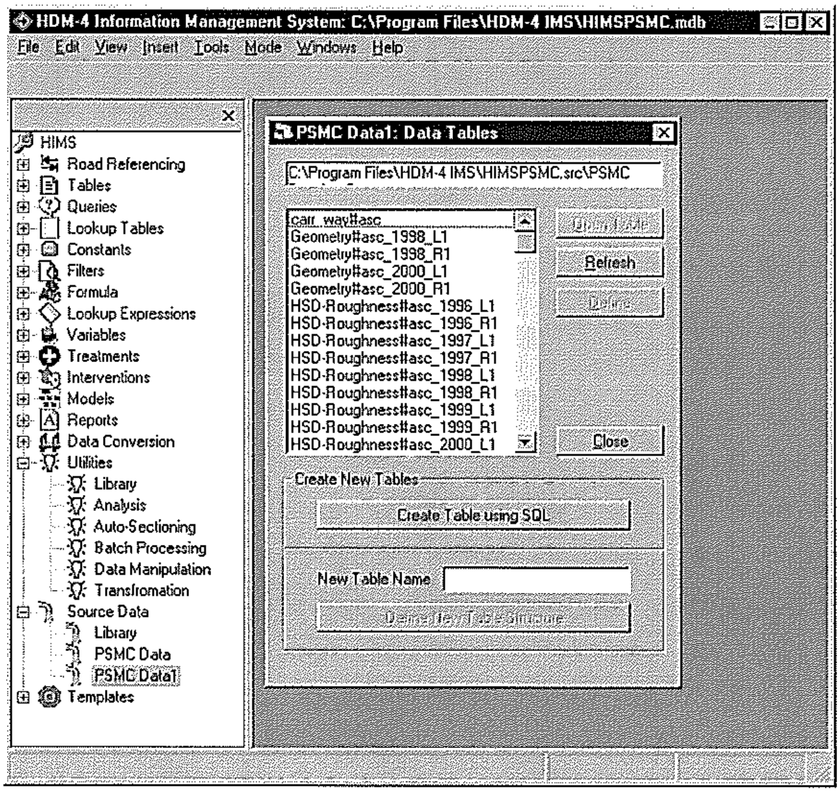


Figure B 13 Organised Data Tables.

❑ Create 100 m sections

In HIMS, add another source database called PSMC Data 100 as described above. Create the 100 m sectioning table in this database and then transform the data into this database.

Select **Tools|Utilities|Auto-Sectioning** to start the automatic sectioning wizard. The following steps are involved:

- Define Data Source: use PSMC Data1.mdb, and use the Carriage way table as the source table (see Figure B 14). This will include all the roads inside that network. If data for only one road is supplied, choose any other source data table. Select Next to continue after source data information is defined.

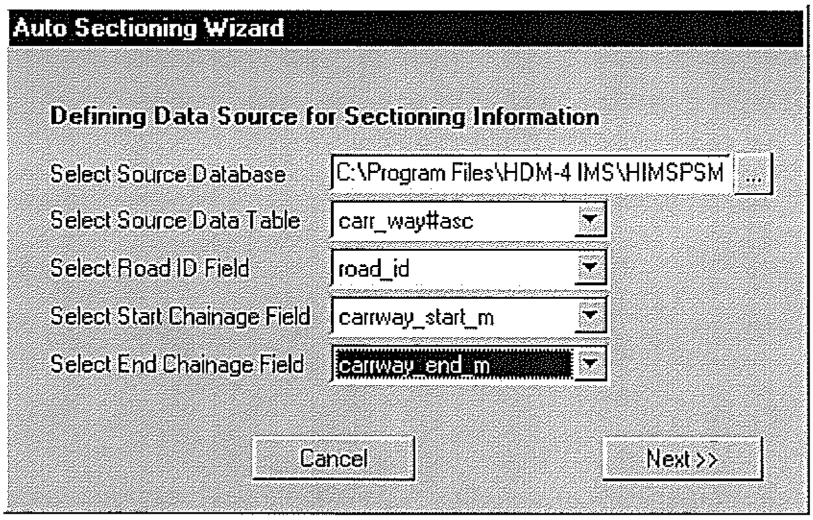


Figure B.14 Define Data Source for Sectioning.

- Define Sectioning Method: here we use the first option: create section based on road length (see Figure B.15). Then select Next.

- Defining Chainage Unit and Interval: use the default (see Figure B 16). Select Next.

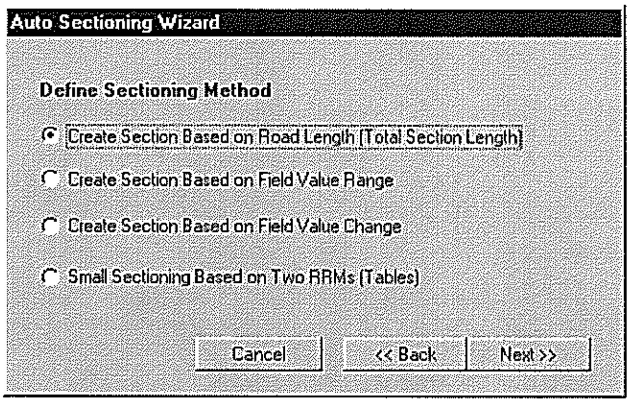


Figure B.15 Define Sectioning Method

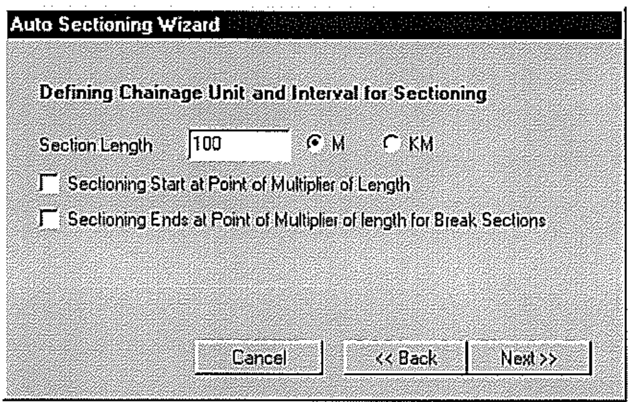


Figure B 16 Defining Chainage Unit and Interval for Sectioning.

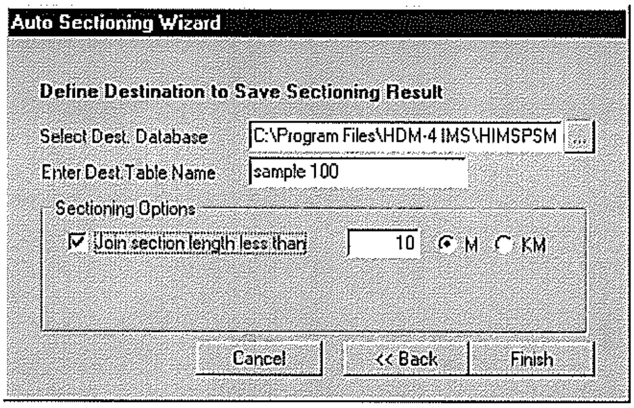


Figure B 17 Define Destination for Sectioning.

- Define Destination for Sectioning: we select PSMC Data 100.mdb as the database, use sample 100 as the table name, and click the checkbox for Joining section length less than 10 m (see Figure B 17). Select Finish to begin the automatic sectioning. The table Sample 100 will be used as the reference to transform data.

- **Modify Field Name of Sample 100 table:**

The sample 100 table uses the same field names as in the carriageway table so they need to be changed. Select PSMC Data 100 in the HIMS management panel, then select Display Data Tables to display the tables (see Figure B 18).

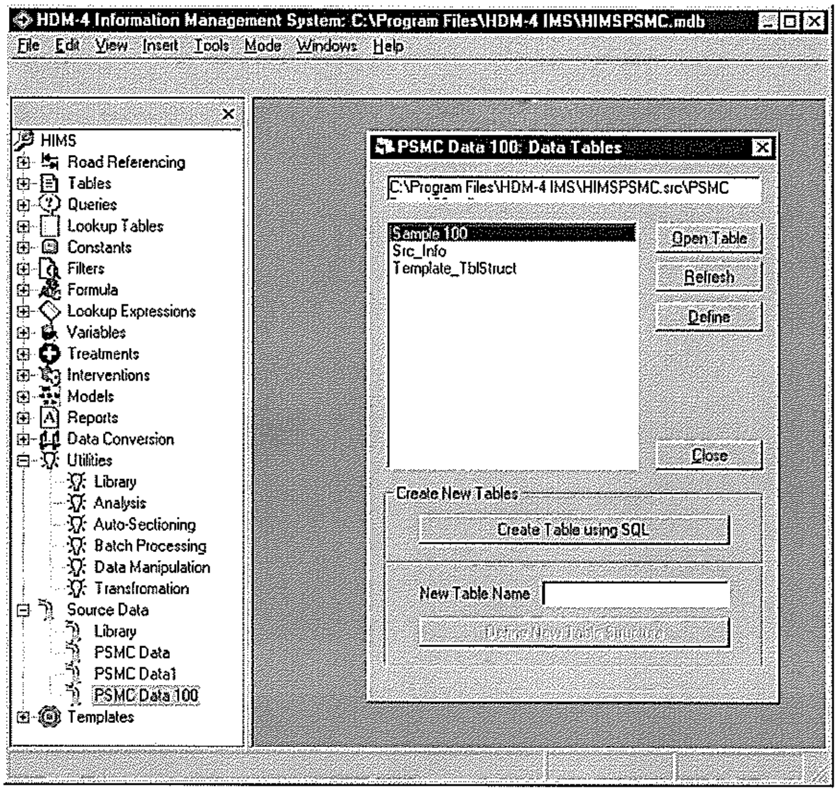


Figure B 18 Sample 100 in PSMC Data 100.

Select Define to open the table structure to modify the field names (change Carrway_Start_m to Start_m, Carrway_End_m to End_m, see Figure B 19).

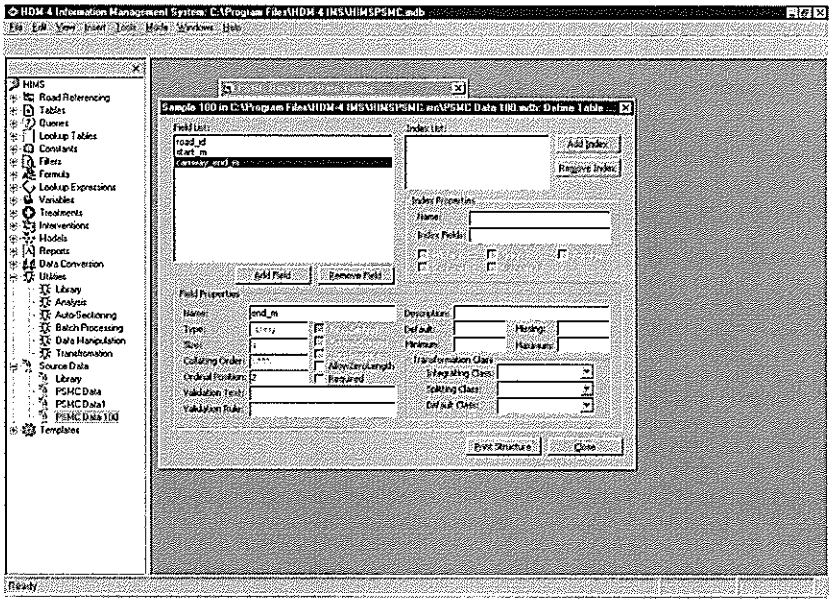


Figure B 19 Change the Field Name in Sample 100.

❑ Define Data Transformation Batch

The data transformations should be done in a batch for all the tables. This batch includes two groups: transform surface/pavement data into other tables inside the same PSMC Data 1 database, and then transform all tables (not including treatment table) into PSMC Data 100 according to the section definition in Sample 100 table.
The batch definition includes:

- o Defining field transformation class
- o Defining data transformation batch

Transformation class is used to define the way to process data when a transformation happens. There are three transformation classes:

- o Integrating class: defines the way data are processed when small sections are combined into a large section.
- o Splitting class: defines the way data are processed when a big section is split into several small sections.
- o Default Class: define the way data are obtained when a missed section is found.

Depending on the data type of a field, different options are available for different classes. Please refer to the user manual and technical reference of HIMS software.

The following are the detailed descriptions.

- **Defining field transformation class**. This includes selecting the fields to be transformed, and the transformation class for each field.

  The following transformation class definition tables are defined using HIMS Transformation Wizard:

  - Trans1: from surface table to any other data table of the same year, or from right surface table to rating table of the same year
  - Trans11: from surface table to any other data table of next year, or from right surface table to rating table of the next year
  - Trans2: from pavement table to any other data table of the same year, or from right pavement table to the rating table of the same year
  - Trans21: from pavement table to any other data table of the next year, or from right pavement table to the rating table of next year
  - Trans3: from left surface table to rating table of the same year
  - Trans31: from left surface table to rating table of the next year
  - Trans4: from left pavement table to rating table of the same year
  - Trans41: from left pavement table to rating table of the next year
  - Trans5: from left surface table to rating table of the same year while the inspection start_m and end_m are used instead
  - Trans51: from left surface table to rating table of the next year while inspection strat_m and end_m are used instead
  - Trans6: from right surface table to rating table of the same year while inspection start_m and end_m are used instead
  - Trans61: from right surface table to rating table of the next year while inspection start_m and end_m are used instead
  - Trans7: from left pavement table to rating table of the same year while inspection start_m and end_m are used instead
  - Trans71: from left pavement table to rating table of the next year while inspection start_m and end_m are used instead
  - Trans8: from right pavement table to rating table of the same year while inspection start_m and end_m are used instead
  - Trans81: from right pavement table to rating table of the next year while inspection start_m and end_m are used instead
  - TransGeo: from HSD-Geometry tables in PSMC Data1 to HSD-Geometry (100m s) tables in PSMC Data 100 TransHSDR: from HSD-Roughness tables in PSMC Data1 to HSD-Roughness (100m s) tables in PSMC Data 100
  - TransHSDRu: from HSD-Rutting tables in PSMC Data1 to HSD-Rutting (100m s) tables in PSMC Data 100
  - TransHSDS: from HSD-Skid tables in PSMC Data1 to HSD-Skid (100m s) tables in PSMC Data 100
  - TransHSDT: from HSD-Texturetables in PSMC Data1 to HSD-Texture (100m s) tables in PSMC Data 100

- TransR: from Rating tables in PSMC Data1 to Rating (100m s) tables in PSMC Data 100
- TransRo: from Roughness tables in PSMC Data1 to Roughness (100m s) tables in PSMC Data 100

To define field transformation class table, select **Tools|Utilities|Transformation** to open the data transformation wizard. This wizard is used to define the field transformation class. Simply select PSMC Data1 as the source database, and then select a table to define the transformation class to be used. For each type of data, only one transformation definition table is needed, e.g. for all Geometry data only one transformation class definition will be used. Figure B.20 is an example of defining the TransGeo table.

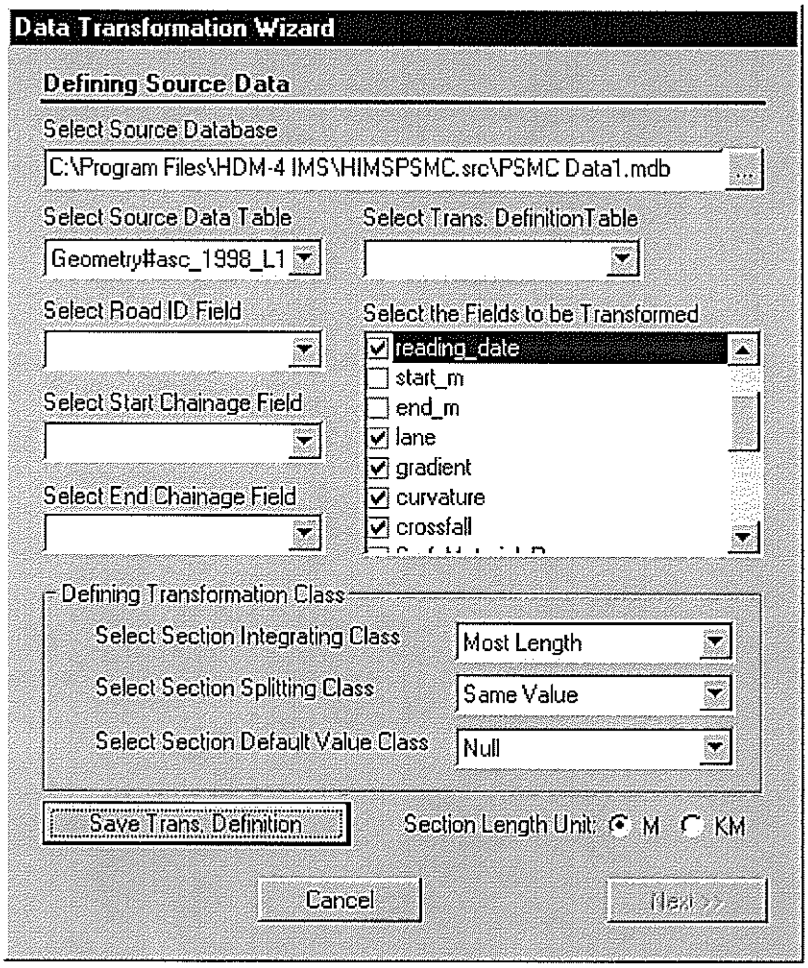


Figure B 20 Defining Table Transformation Definition.

Select a Geometry table, e.g. Geometry#asc_1998_L1, and then select those fields to be transformed to the destination table (e.g. Reading_Date, Lane, Gradient, Crossfall, Curvature), and define the transformation class for each field, and then select Save Trans. Definition to save it as TransGeo. Note that you can also go to second step of the transformation wizard to define the target name for each field. In this case define the Road_ID, Start and End field to activate the Next button. Be sure to click Back, rather than click Finish which runs the transformation, and then select a different source data table to define and then save as a different name. All the definition tables are saved inside the source database.

Please note: defining the transformation definition in this way is cumbersome since every field must exist before one can select and define it. Another way to do it is to open the source database in Access (or in HIMS, as long as in table format), and simply copy an existing template of the definition table, then add new records to it. Any field name can be added as long as the field will be created before it is involved in a transformation.

- **Defining data transformation batch**

  Select **Tools|Utilities|Batch Processing** and the following screen will appear (see Figure B 21). Select **Edit** to add new batch processing definition or edit an existing one (see Figure B 22). Currently only one Batch Processing table can be used and stored inside a database Transformation.mdb under the system directory. For other batch processing functions please see the HIMS user manual.

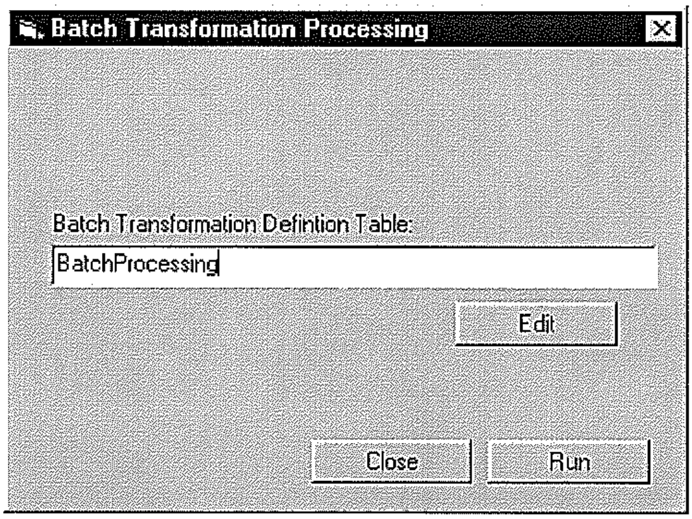


Figure B 21 Batch Transformation Processing.

In Figure B 22, select Define to edit or add a record, this will automatically invoke the Transformation Wizard for definition. The transformation is done as described in the Transform Carriageway width to data tables discussed earlier.

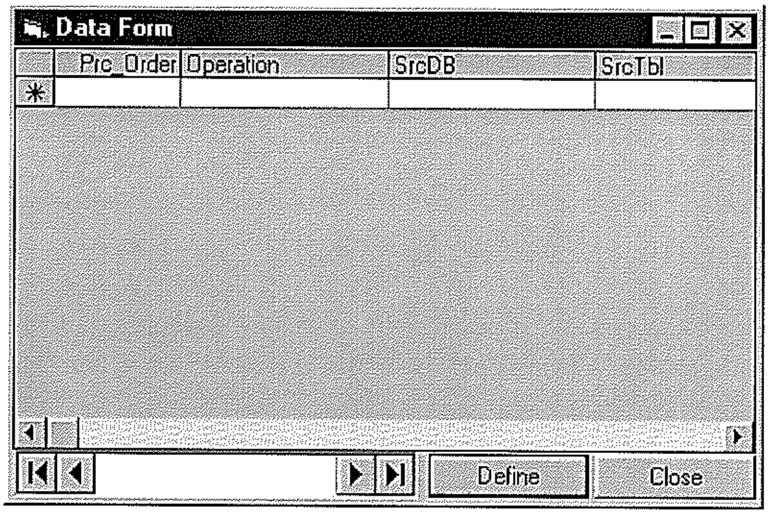


Figure B.22 Data Form for Defining Batch Processing.

The following is an example to define Geometry#asc_1998_L1 transformation:

After selecting Define in Figure B22, the Transformation Wizard will appear (see Figure B23). Select the data source properties, and be sure to select the Trans. Definition table. Then select Next to define destination data properties (see Figure B24). Then select Section Def. Table instead of Destination table, and use the Sample 100 table for creating new tables. Select Finish to close the Wizard. A new record will be added (see Figure B25).

Please note, all the transformations can be defined in different orders as long as the correct order is achieved at completion by modifying the value in Prc_Order column (see Figure B25).

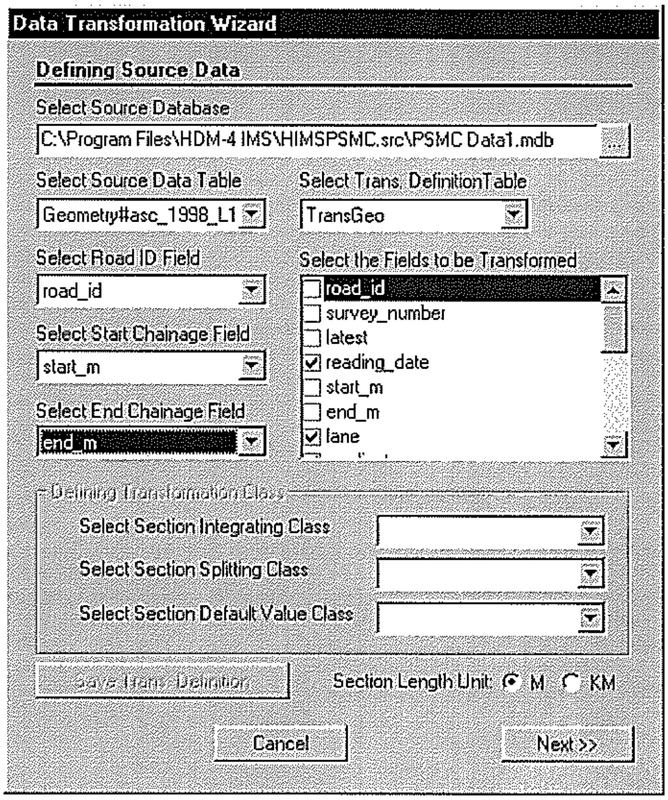


Figure B 23 Define Data Source Properties.

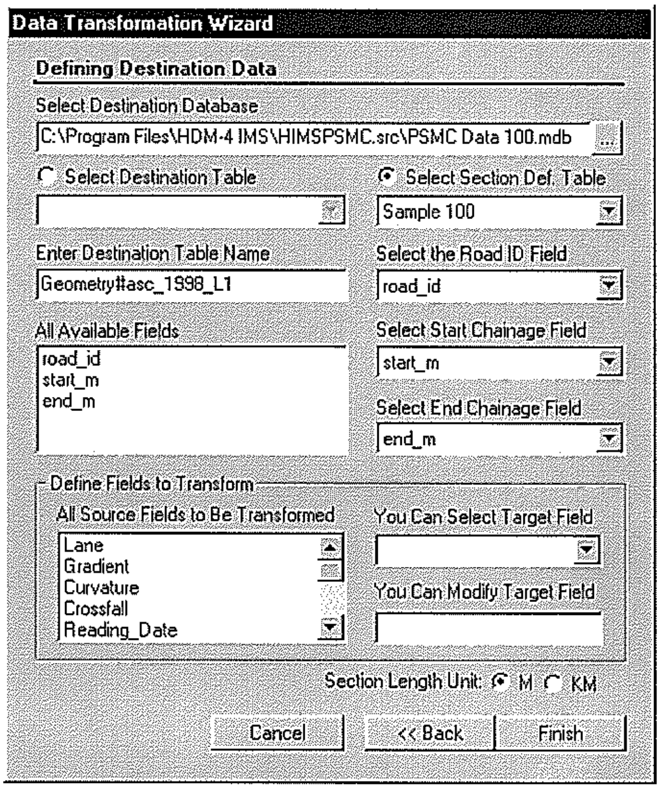
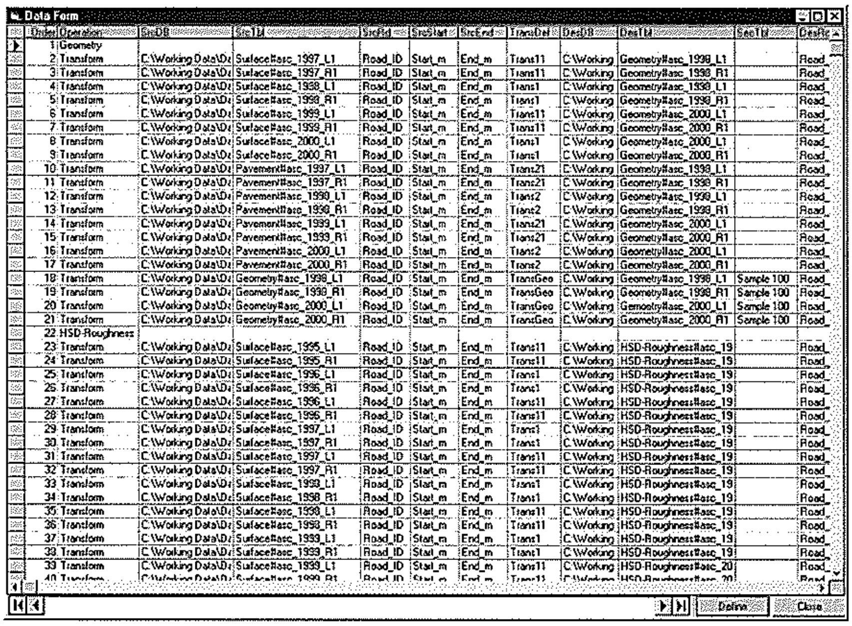


Figure B.24 Define Destination Data Properties.



Data Form

Order	Operation	SrcDB	SrcTbl	SrcFld	SrcStart	SrcEnd	TransDef	DesDB	DesTbl	SecTbl	DesFlc
1	Geometry										
2	Transform	C:\Working Data\Da	Surface#asc_1997_L1	Road_ID	Start_m	End_m	Trans11	C:\Working	Geometry#asc_1998_L1		Road_
3	Transform	C:\Working Data\Da	Surface#asc_1997_R1	Road_ID	Start_m	End_m	Trans11	C:\Working	Geometry#asc_1998_R1		Road_
4	Transform	C:\Working Data\Da	Surface#asc_1998_L1	Road_ID	Start_m	End_m	Trans1	C:\Working	Geometry#asc_1998_L1		Road_
5	Transform	C:\Working Data\Da	Surface#asc_1998_R1	Road_ID	Start_m	End_m	Trans1	C:\Working	Geometry#asc_1998_R1		Road_
6	Transform	C:\Working Data\Da	Surface#asc_1999_L1	Road_ID	Start_m	End_m	Trans11	C:\Working	Geometry#asc_2000_L1		Road_
7	Transform	C:\Working Data\Da	Surface#asc_1999_R1	Road_ID	Start_m	End_m	Trans11	C:\Working	Geometry#asc_2000_R1		Road_
8	Transform	C:\Working Data\Da	Surface#asc_2000_L1	Road_ID	Start_m	End_m	Trans1	C:\Working	Geometry#asc_2000_L1		Road_
9	Transform	C:\Working Data\Da	Surface#asc_2000_R1	Road_ID	Start_m	End_m	Trans1	C:\Working	Geometry#asc_2000_R1		Road_
10	Transform	C:\Working Data\Da	Pavement#asc_1997_L1	Road_ID	Start_m	End_m	Trans21	C:\Working	Geometry#asc_1998_L1		Road_
11	Transform	C:\Working Data\Da	Pavement#asc_1997_R1	Road_ID	Start_m	End_m	Trans21	C:\Working	Geometry#asc_1998_R1		Road_
12	Transform	C:\Working Data\Da	Pavement#asc_1998_L1	Road_ID	Start_m	End_m	Trans2	C:\Working	Geometry#asc_1998_L1		Road_
13	Transform	C:\Working Data\Da	Pavement#asc_1998_R1	Road_ID	Start_m	End_m	Trans2	C:\Working	Geometry#asc_1998_R1		Road_
14	Transform	C:\Working Data\Da	Pavement#asc_1999_L1	Road_ID	Start_m	End_m	Trans21	C:\Working	Geometry#asc_2000_L1		Road_
15	Transform	C:\Working Data\Da	Pavement#asc_1999_R1	Road_ID	Start_m	End_m	Trans21	C:\Working	Geometry#asc_2000_R1		Road_
16	Transform	C:\Working Data\Da	Pavement#asc_2000_L1	Road_ID	Start_m	End_m	Trans2	C:\Working	Geometry#asc_2000_L1		Road_
17	Transform	C:\Working Data\Da	Pavement#asc_2000_R1	Road_ID	Start_m	End_m	Trans2	C:\Working	Geometry#asc_2000_R1		Road_
18	Transform	C:\Working Data\Da	Geometry#asc_1998_L1	Road_ID	Start_m	End_m	TransGeo	C:\Working	Geometry#asc_1998_L1	Sample 100	Road_
19	Transform	C:\Working Data\Da	Geometry#asc_1998_R1	Road_ID	Start_m	End_m	TransGeo	C:\Working	Geometry#asc_1998_R1	Sample 100	Road_
20	Transform	C:\Working Data\Da	Geometry#asc_2000_L1	Road_ID	Start_m	End_m	TransGeo	C:\Working	Gemoetry#asc_2000_L1	Sample 100	Road_
21	Transform	C:\Working Data\Da	Geometry#asc_2000_R1	Road_ID	Start_m	End_m	TransGeo	C:\Working	Geometry#asc_2000_R1	Sample 100	Road_
22	HSD-Roughness										
23	Transform	C:\Working Data\Da	Surface#asc_1995_L1	Road_ID	Start_m	End_m	Trans11	C:\Working	HSD-Roughness#asc_19		Road_
24	Transform	C:\Working Data\Da	Surface#asc_1995_R1	Road_ID	Start_m	End_m	Trans11	C:\Working	HSD-Roughness#asc_19		Road_
25	Transform	C:\Working Data\Da	Surface#asc_1996_L1	Road_ID	Start_m	End_m	Trans1	C:\Working	HSD-Roughness#asc_19		Road_
26	Transform	C:\Working Data\Da	Surface#asc_1996_R1	Road_ID	Start_m	End_m	Trans1	C:\Working	HSD-Roughness#asc_19		Road_
27	Transform	C:\Working Data\Da	Surface#asc_1996_L1	Road_ID	Start_m	End_m	Trans11	C:\Working	HSD-Roughness#asc_19		Road_
28	Transform	C:\Working Data\Da	Surface#asc_1996_R1	Road_ID	Start_m	End_m	Trans11	C:\Working	HSD-Roughness#asc_19		Road_
29	Transform	C:\Working Data\Da	Surface#asc_1997_L1	Road_ID	Start_m	End_m	Trans1	C:\Working	HSD-Roughness#asc_19		Road_
30	Transform	C:\Working Data\Da	Surface#asc_1997_R1	Road_ID	Start_m	End_m	Trans1	C:\Working	HSD-Roughness#asc_19		Road_
31	Transform	C:\Working Data\Da	Surface#asc_1997_L1	Road_ID	Start_m	End_m	Trans11	C:\Working	HSD-Roughness#asc_19		Road_
32	Transform	C:\Working Data\Da	Surface#asc_1997_R1	Road_ID	Start_m	End_m	Trans11	C:\Working	HSD-Roughness#asc_19		Road_
33	Transform	C:\Working Data\Da	Surface#asc_1998_L1	Road_ID	Start_m	End_m	Trans1	C:\Working	HSD-Roughness#asc_19		Road_
34	Transform	C:\Working Data\Da	Surface#asc_1998_R1	Road_ID	Start_m	End_m	Trans1	C:\Working	HSD-Roughness#asc_19		Road_
35	Transform	C:\Working Data\Da	Surface#asc_1998_L1	Road_ID	Start_m	End_m	Trans11	C:\Working	HSD-Roughness#asc_19		Road_
36	Transform	C:\Working Data\Da	Surface#asc_1998_R1	Road_ID	Start_m	End_m	Trans11	C:\Working	HSD-Roughness#asc_19		Road_
37	Transform	C:\Working Data\Da	Surface#asc_1999_L1	Road_ID	Start_m	End_m	Trans1	C:\Working	HSD-Roughness#asc_19		Road_
38	Transform	C:\Working Data\Da	Surface#asc_1999_R1	Road_ID	Start_m	End_m	Trans1	C:\Working	HSD-Roughness#asc_19		Road_
39	Transform	C:\Working Data\Da	Surface#asc_1999_L1	Road_ID	Start_m	End_m	Trans11	C:\Working	HSD-Roughness#asc_20		Road_
40	Transform	C:\Working Data\Da	Surface#asc_1999_R1	Road_ID	Start_m	End_m	Trans11	C:\Working	HSD-Roughness#asc_20		Road_

Define Close

Figure B 25 A Transformation Definition .

The transformation batch is directly related to the actual number of tables. If a table is there, then one can define a record, if it is not there, it will not need to be defined. The basic rule for defining the batch record here is as follows:

- For treatment data, define transformation of the same year, and the year after if the data exists, as follows:

  surface#asc_L_1999 → geometry#asc_L1_1999
  surface#asc_L_1999 → geometry#asc_L1_2000
  surface#asc_L_1999 → rating#asc_1999
  surface#asc_L_1999 →rating#asc_2000
  pavement#asc_L_1999 →geometry#asc_L1_1999
  pavement#asc_L_1999 →geometry#asc_L1_2000
  pavement#asc_L_1999 →rating#asc_L1_1999
  pavement#asc_L_1999 →rating#asc_L1_2000

The reason for this is that a survey may happen before or after a treatment. If the survey happened before the treatment, the treatment was considered as having happened in the following year; if the survey happened after the treatment, then the treatment was considered to have happened in the current year.

Defining batch transformation in this manner is time consuming, but it reduces the likelihood of making a mistake. An alternative way is to open the database in Access, and input and edit the table. The table can be found in C:\Windows\System\Transformation.mdb file. Another advantage of doing it this way is that even if a table does not yet exist before the batch is run, you still can define it as long as it will exist before that record is activated to run.

Please note, depending on the location to run the whole process, file location in the batch processing table will be different. The HIMS or Access can be used to replace the location definition.

Once the definition is finished, a copy should be made for checking and modifying later if necessary.

- **Run Data Transformation Batch**

As shown earlier in Figure B21, selecting Run started the batch processing. This is very resource and time consuming, so a fast computer with a large amount of hard disk space is recommended. Alternatively, the Transformation Wizard can be used. By double clicking on an empty place on the form, a configuration screen will appear (see Figure B26). If disk space is at a premium, select 'Delete all temporary tables' and 'Compact database after termination'.

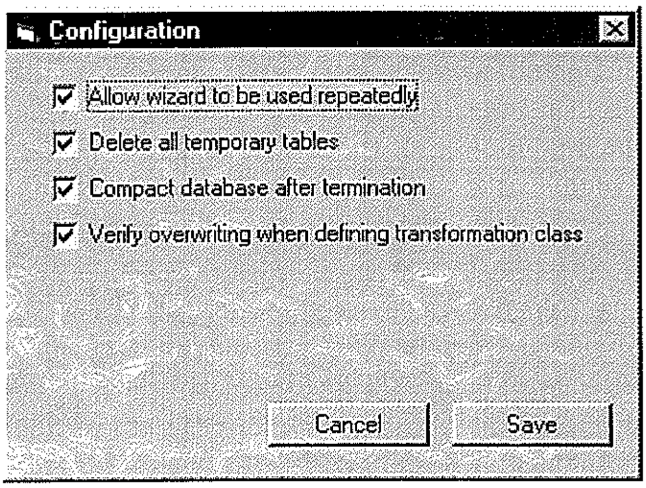


Figure B26 Wizard Configuration.

After the transformation, every table will include information about the treatment and the year it occurs. Data from varying years needs to be organised according to different parameters (e.g. crossfall, NAASRA etc.), and the year of maintenance as described the Database Structure part in Project Database. This includes the following steps:

Create tables for specified parameter
Tidy up the maintenance information
Insert road names into data tables
Import data into a final database

- **Create/tidy up tables for different parameter, and populate road names**

A template database (Sample 100.mdb) was specially created for this project with two forms defined and programmed in Access for tidying up data (see Figure B27).

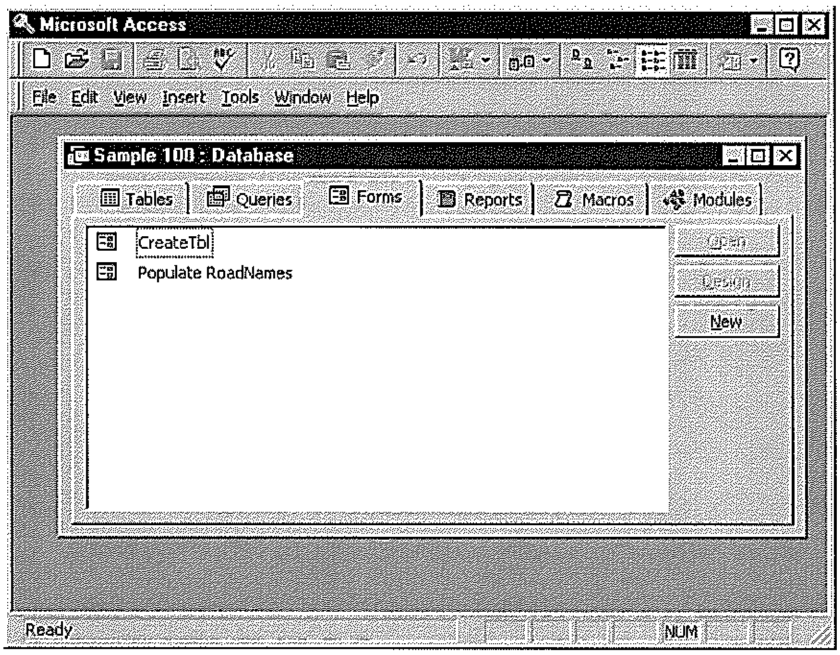


Figure B 27 Template Database For Tidying Up Result.

The first form is for creating and tidying up tables and the second form is for populating road names (see Figure B 28).

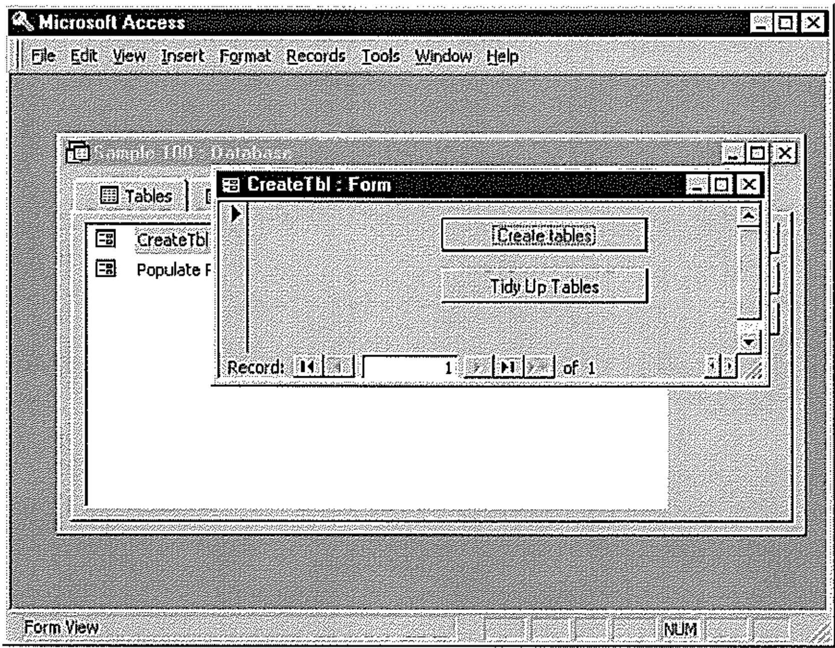


Figure B 28 Form Functions for Creating/Tidying up Tables.

Access is used to create the above. First open database PSMC Data 100.mdb in Access, and then select File|Get External Data|Import to import the two forms. Once imported open the form to run each operation one by one. For example, open CreateTbl form to creating tables; open Tidying up tables; etc.

- **Import data into another final database**

This is done inside HIMS but not inside the HIMS database. Open HIMS, select File|New Empty database to create a new empty database PSMC Data 101.mdb in the same directory as other databases, and then select File|Get External Data to open the data import window (see Figure B 29). Select PSMC Data 100.mdb as the database to export data from, and then select all the tables needed and all the fields needed for each table according to the description in Database structure of Project Database. Mainly all tables without #asc in the name, and all data fields of different year and all maintenance fields.

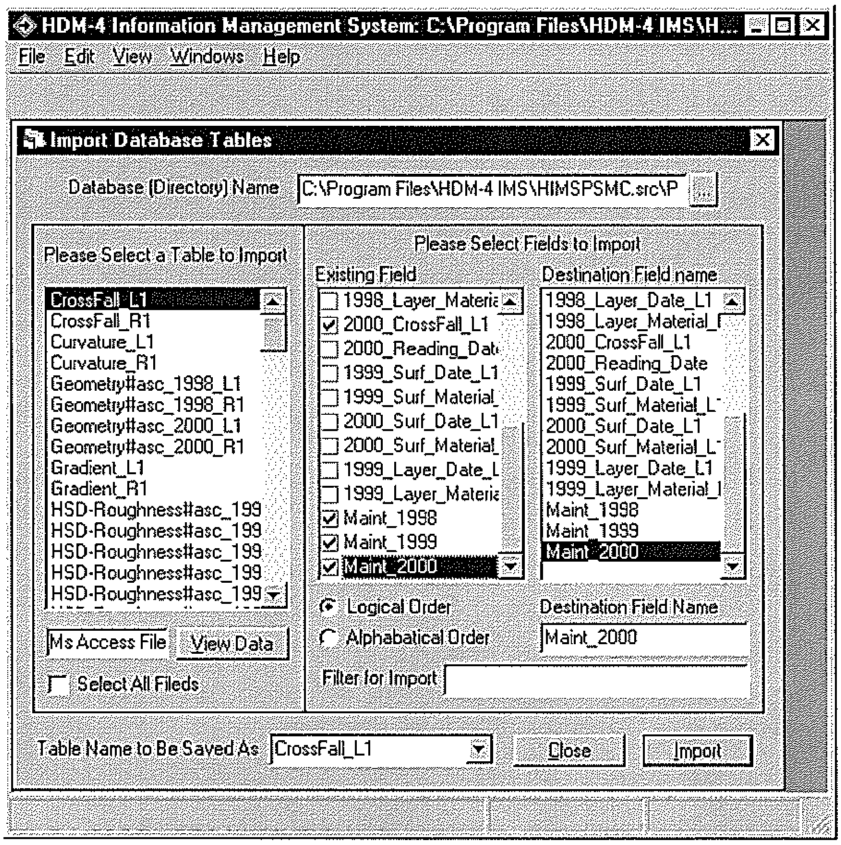


Figure B 29 Import data into PSMC Data 101.

This database PSMC Data 101.mdb will be used for analysis. Please note, however, because not all the necessary source data are available, or RAMM simply does not have the data available, all processes are only based on the actual data, and the processing of data may differ to some degree. Data analysis results may need to be verified with the real road conditions or treatments.

Appendix C. Time Series Database CDs

The time series data CDs have been produced to provide access to the data prepared in the project. Full details of the contents of the CD are given in the file README.TXT which is on CD1. Copies of the CD may be obtained by contacting the author at chris@htc.co.nz.

The CD contains seven data sets:

- ACC Data - Data from Auckland City Council
- NPDC Data - Data from New Plymouth District Council
- NS PSMC Data - North Shore PSMC Data
- NSCC Data - Data from North Shore City Council
- SCC Data - Data from Southland City Council
- TNZ Data1 - Data from Transit NZ (Napier, one road only)
- TNZ Data2 - Data from Transit NZ (Napier, one road only)

To facilitate future analyses similar to this one, the CD contains several templates which can be used with the HIMS to process RAMM data. These include:

- **RAMM Src.mdb**: The standard table structure created using RAMM unloaded road.sql definition file data. This assists with importing into tables inside the database. However, if using data exported from RAMM for Windows, the field names may not match so this must be resolved (although the field name in the source data can be used if it is available). Please note that when pavement structure data is imported, the table pave-structure must be renamed as pavement#asc for further processing.
- **Sample 100**: The standard template for creating 100-m sections and then tidying up the data. Only 3 forms are included with functions attached to the button(s) on the form, they are creating/tidying up tables, populate roadnames (if necessary), and delete empty records.
- **Transfund Template**: Excel files defining source data to be analysed.
- **Transformation Template**: Whole range definition for data transformation, including treatment data to other data and source data to 100-m section data transformation. For a detailed data source, depending on the number of tables existing, transformation Definition may be added, deleted, or modified.

The CDs contain the draft report M001-1-2 01-01-22 Preparation of Databases.doc which describes the procedure used to prepare the databases.

The CDs contain all the files, but those interested in the files suitable for analysis should use:

- ACC Data 101.mdb
- NPDC Data 101.mdb
- PSMC Data 101.mdb
- NSCC Data 101.mdb
- SDC Data 101.mdb
- TNZ Data 101.mdb
- TNZ Data 201.mdb

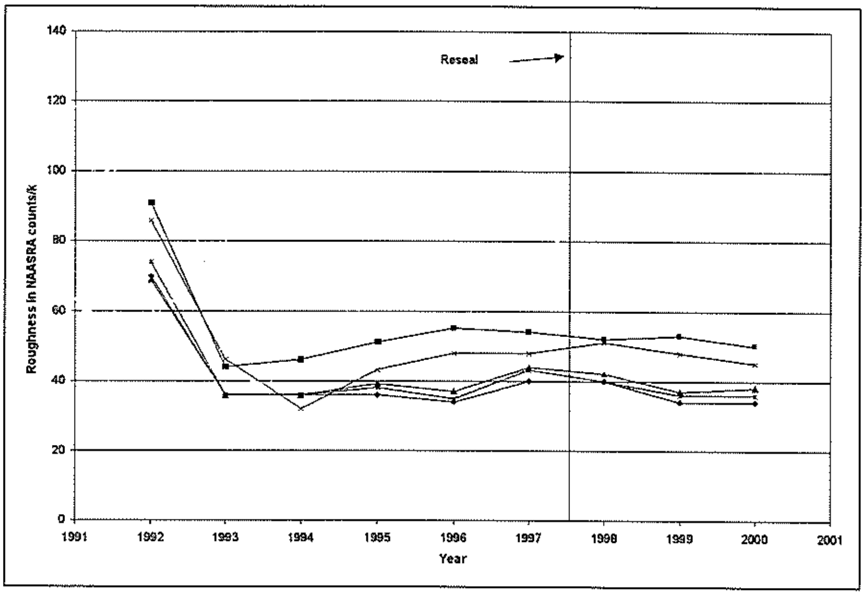
Appendix D. Examples of Roughness Trends

Overview

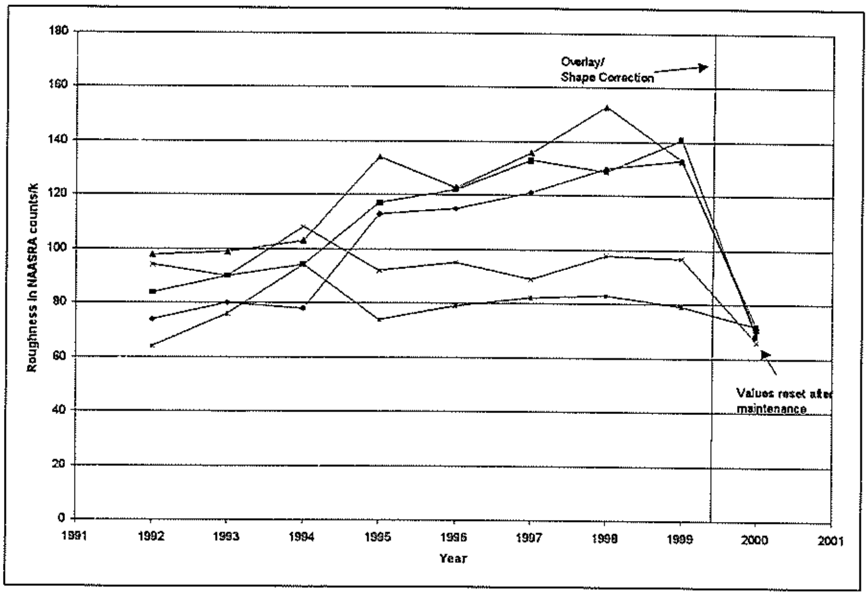
The following figures show typical trends in roughness progression for adjacent 100 m sections on State Highways and a local authority road (i.e. usually 500 m of data). They were extracted at random from the databases.

State Highway Data

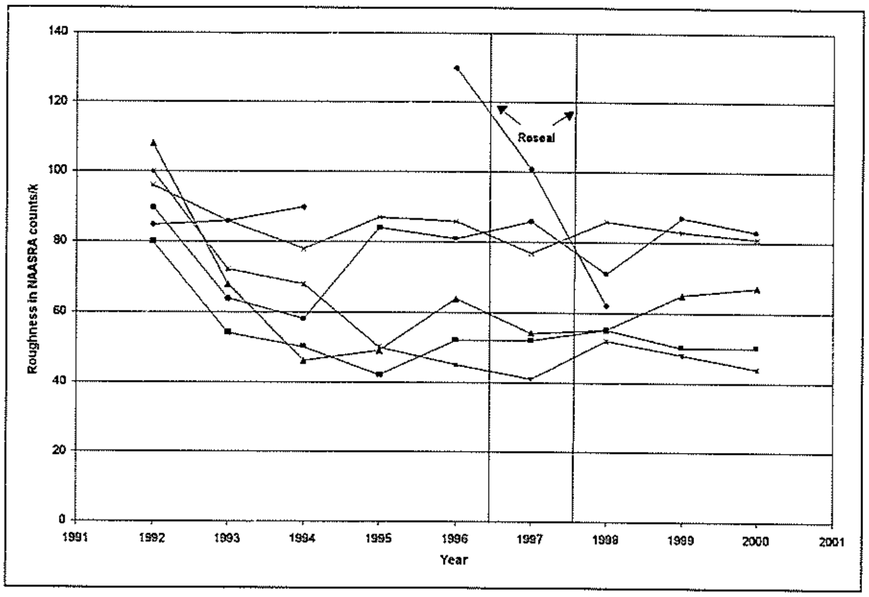
It should be noted that the State Highways prior to 1994 were measured with a response-type meter. Since 1997 the same provider has collected the roughness data with a profilometer.



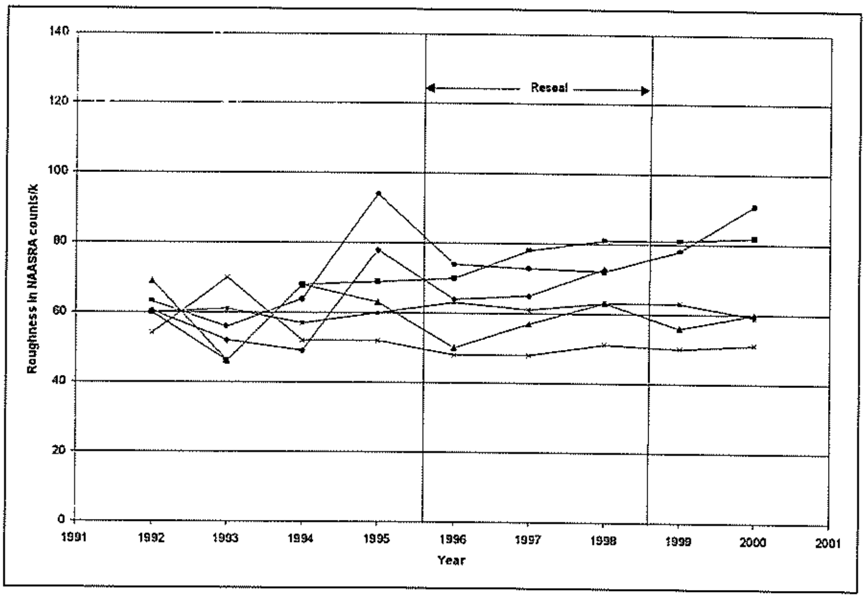
- State Highway data shows good consistency
- The drop in roughness in 1993 suggests that there was a maintenance activity but such a record was absent from the RAMM database
- There was a minor reduction in roughness for most sections after the resurfacing in 1997



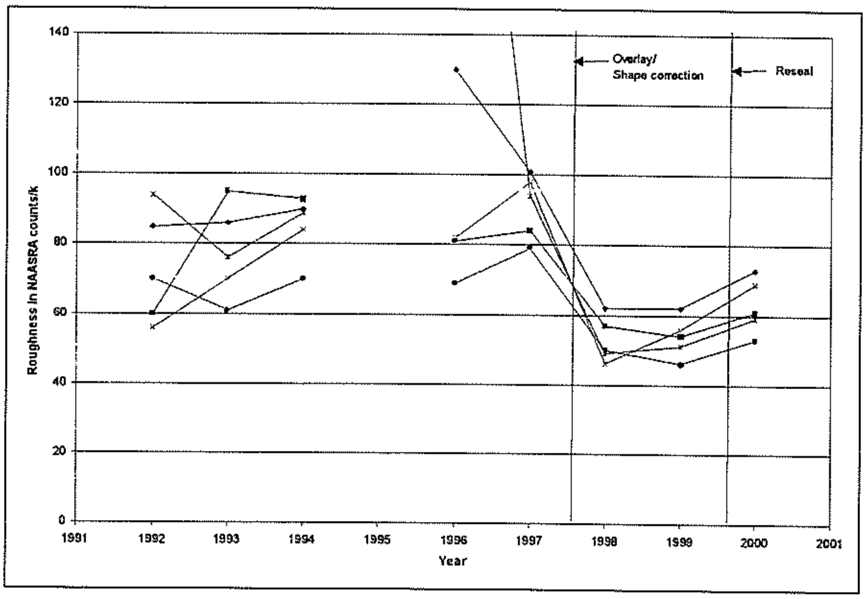
- Data shows some localised variability
- Roughness shows significant reduction after shape correction in 1999



- Data shows high variability
- Minor improvement in some sections after resurfacing in 1996

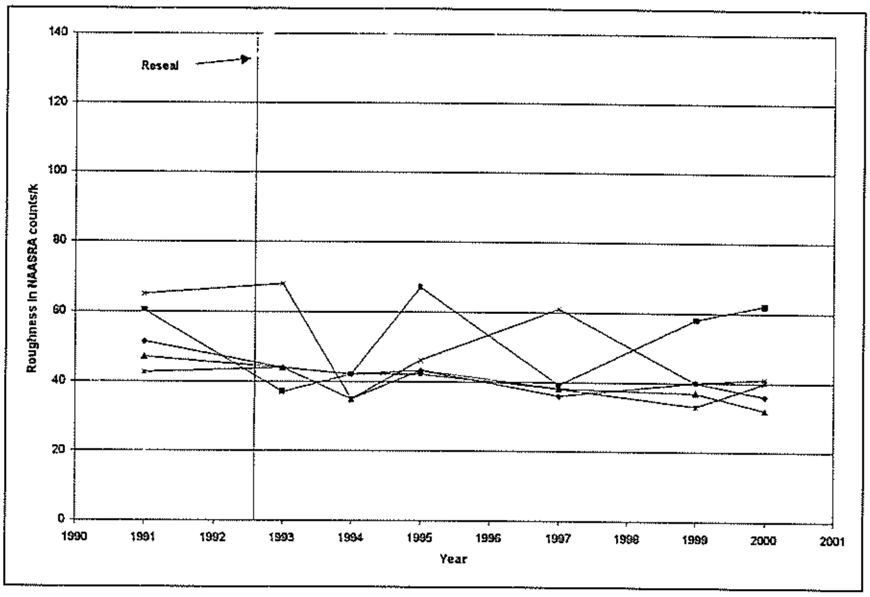


- Data since 1996 shows limited variability
- Resurfacing in 1995 resulted in general decrease in roughness on most sections

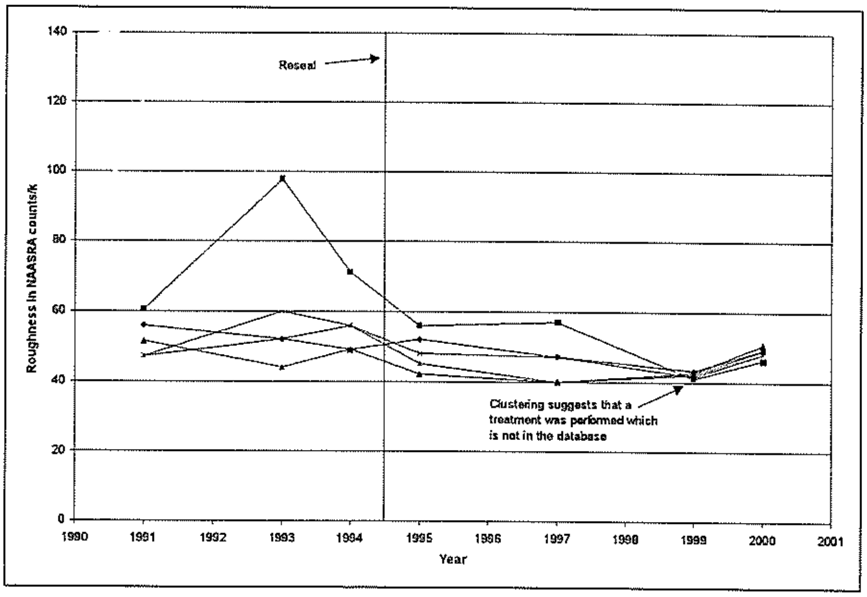


- Data missing for 1995
- Shape correction in 1997 resulted in major decrease in roughness

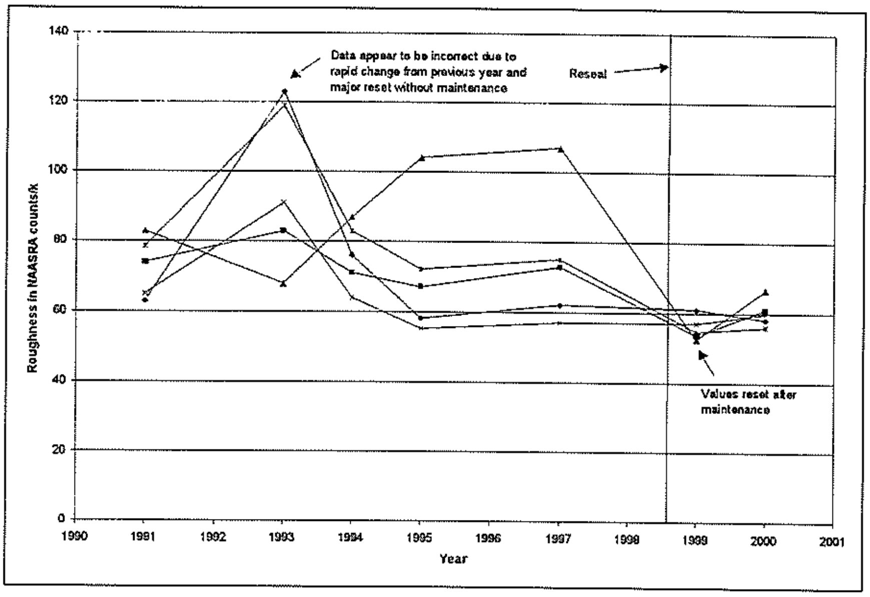
Local Authority Data



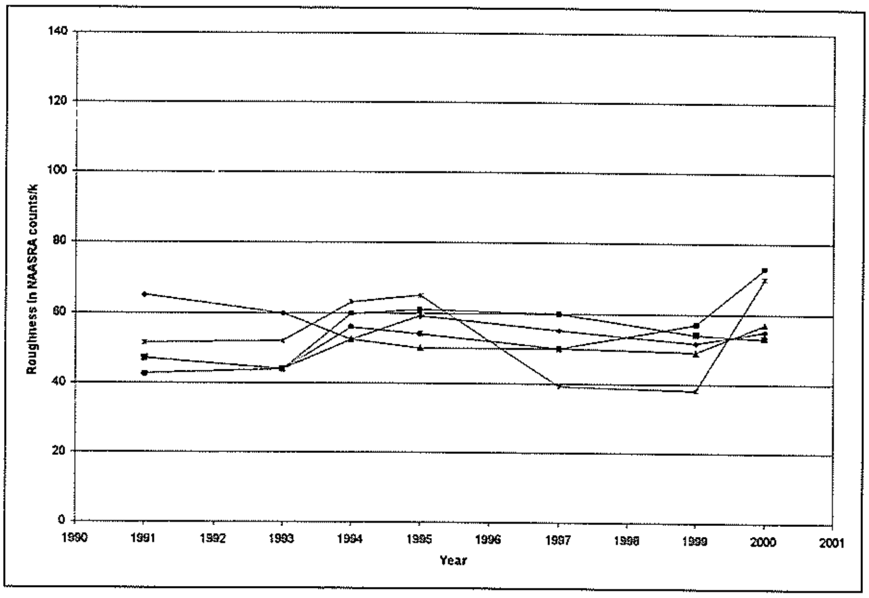
- For some 100-m sections data shows great variability, for others little
- On most sections, resurfacing in 1992 had minor impact



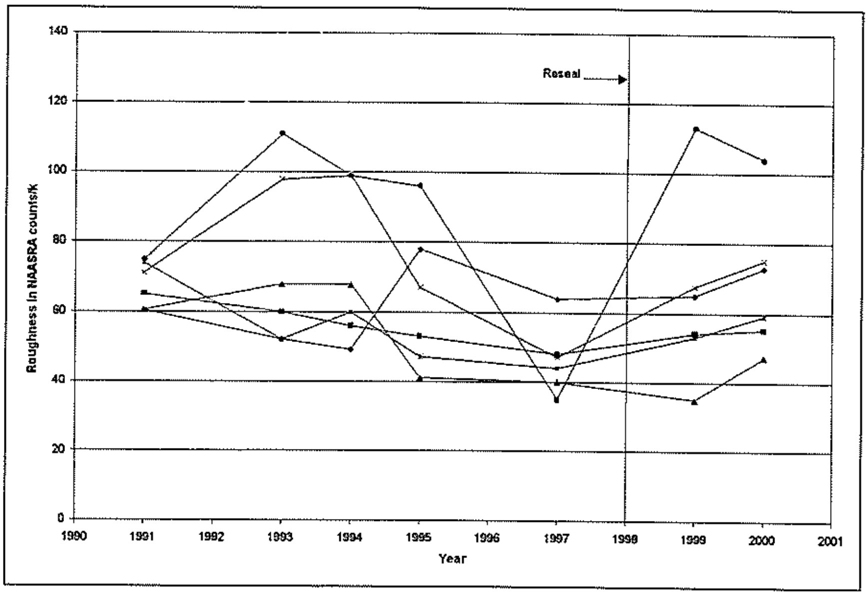
- In 1993 one section had a clear outlier, inconsistent with other data
- Resurfacing caused a decrease in roughness in 1994
- Clustering in 1999 suggests maintenance treatment that was not recorded in RAMM



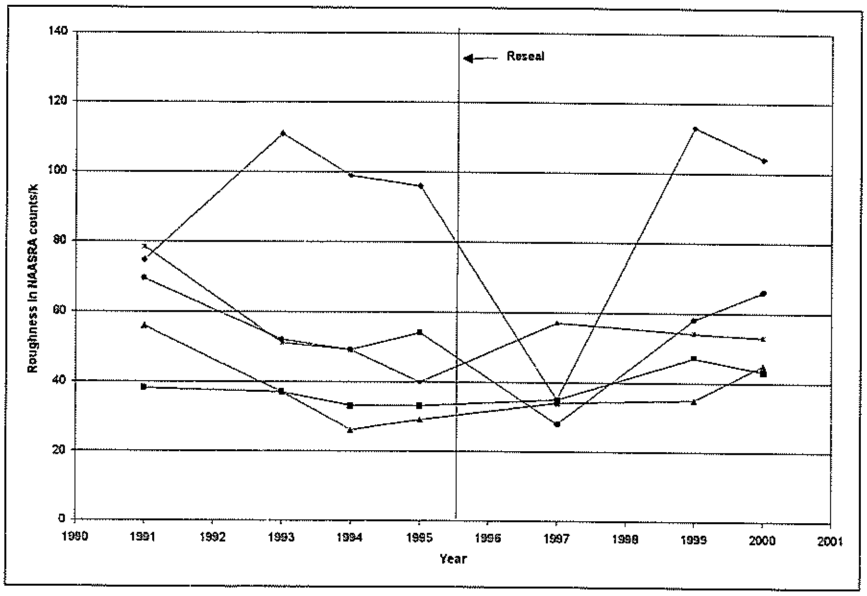
- Some outliers in 1993
- Resurfacing in 1998 resulted in major decrease in roughness for some sections, little impact for those which already had low roughnesses.



- Data shows moderate variability



- Data shows high variability
- Maintenance did not show a major impact on condition



- Data shows high variability
- Maintenance had some impact on some sections

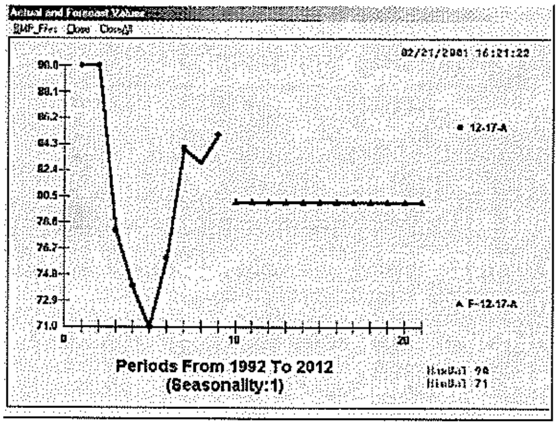
Appendix E. Autobox Time Series Results from Sample Roughness Data

This appendix presents the results for each of the 20 x 100 m sections analysed using the test roughness data. The roughness trend is given along with the best forecast from Autobox.

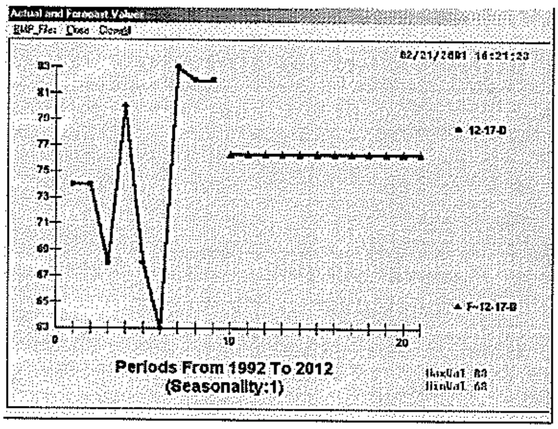
It should be noted that the data set was specifically designed to test Autobox's ability to identify pulses and level shifts, so the failure to establish a reliable time series should not necessarily be viewed as a deficiency in Autobox.

In almost all instances, the forecast from Autobox is a horizontal line: the mean. This is because the input data are inadequate for developing a predictive model and so the best forward estimator is the mean of the existing data.

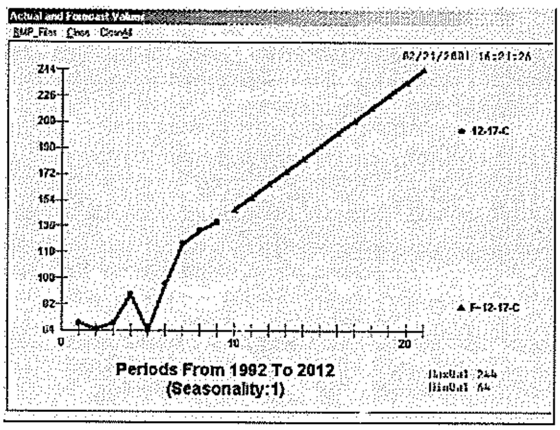
State Highway Data



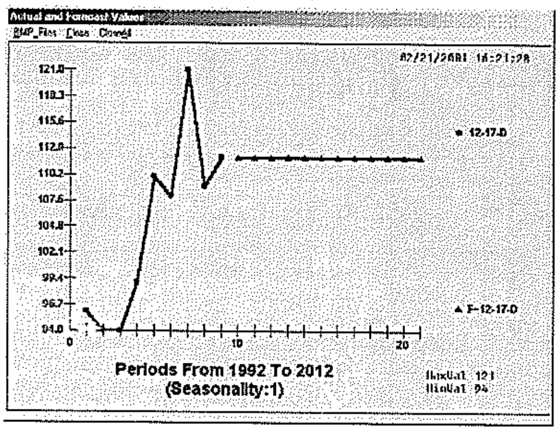
12-17-A



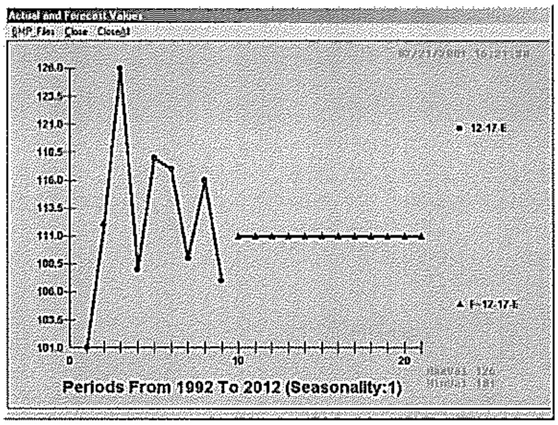
12-17-B



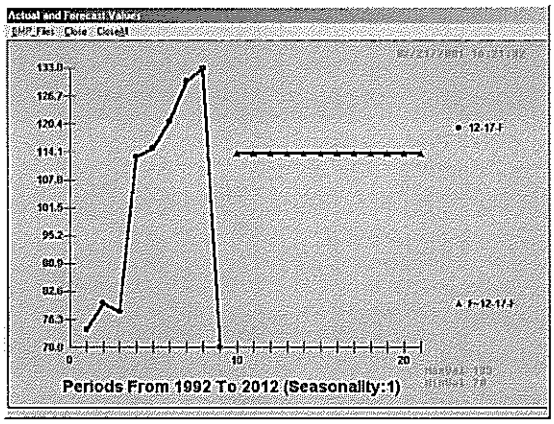
12-17-C



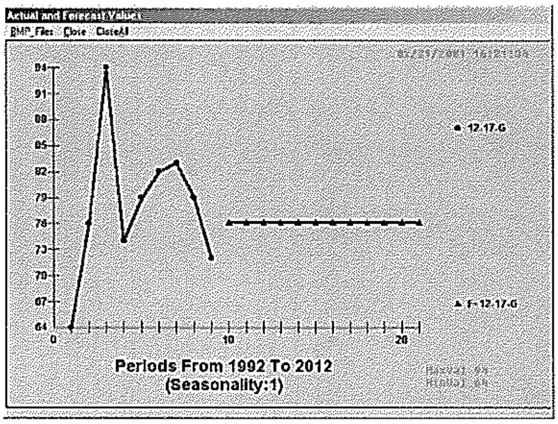
12-17-D



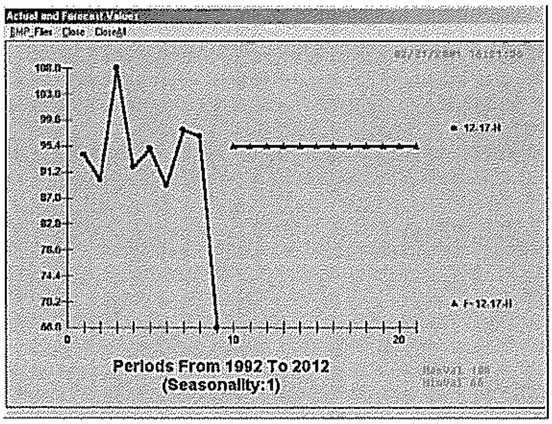
12-17-E



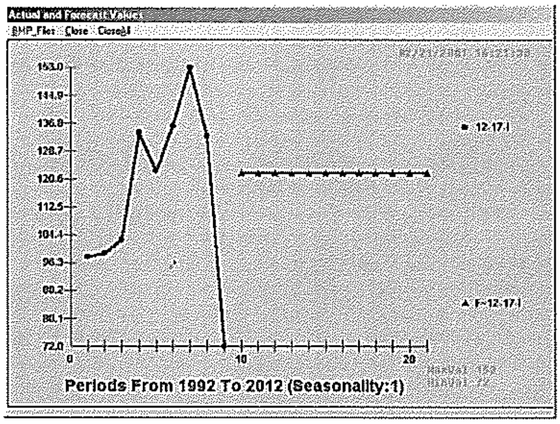
12-17-F



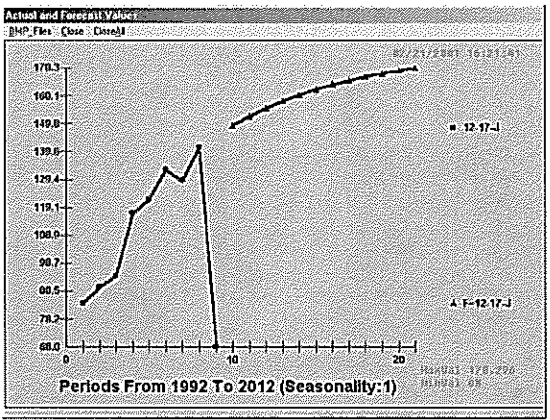
12-17-G



12-17-H

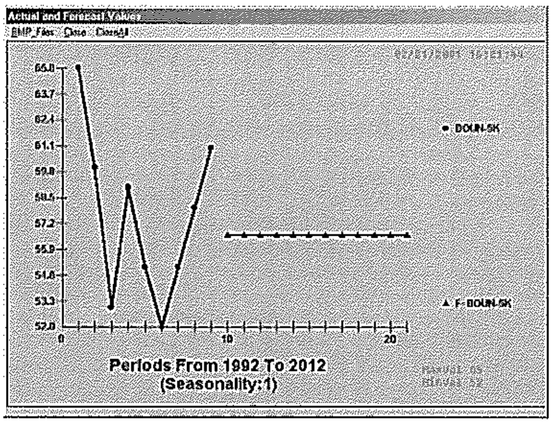


12-17-I

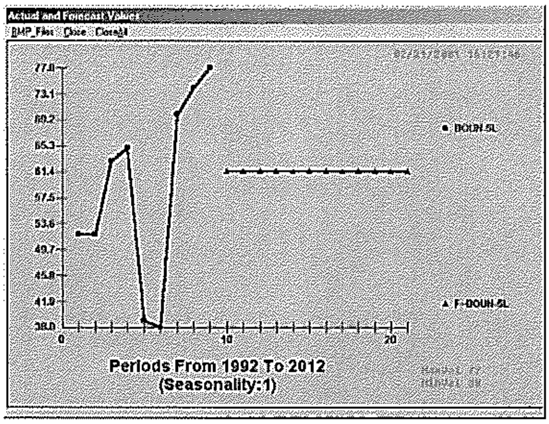


12-17-J

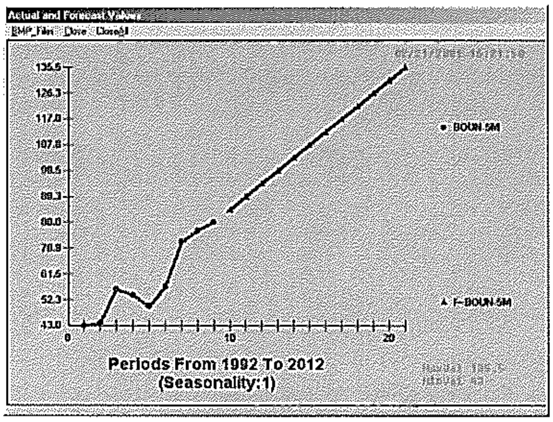
Local Authority Data



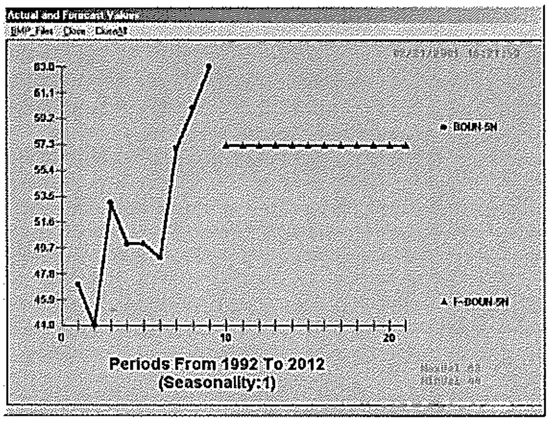
BOUN-5K



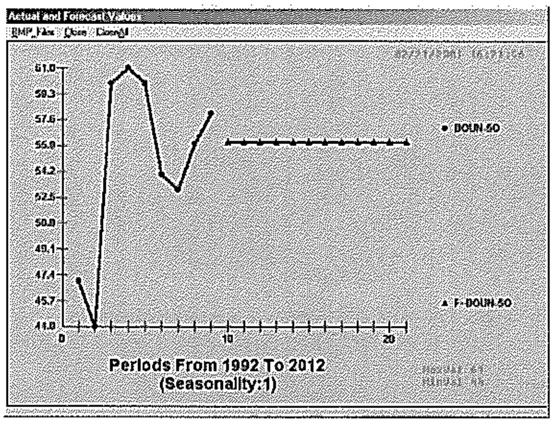
BOUN-5L



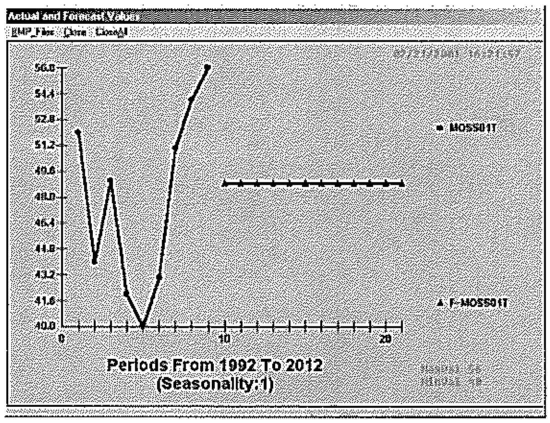
BOUN-5M



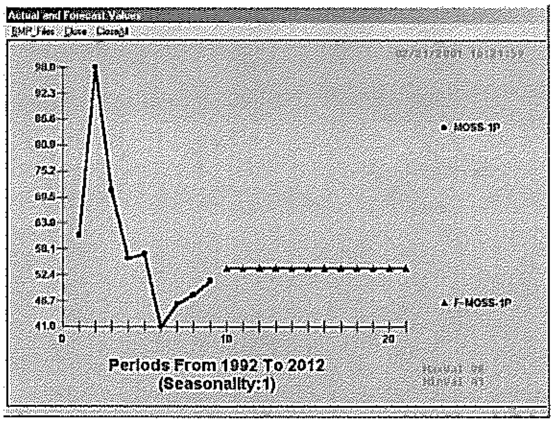
BOUN-5N



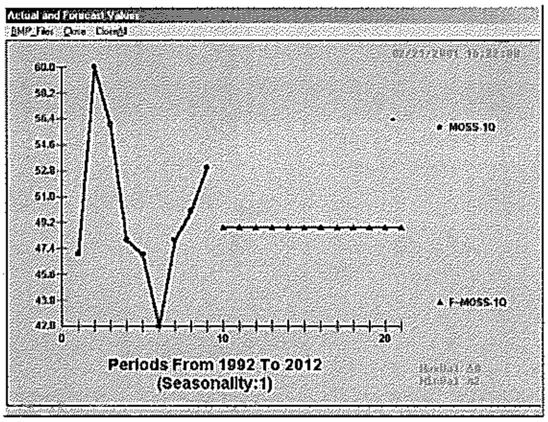
BOUN-50



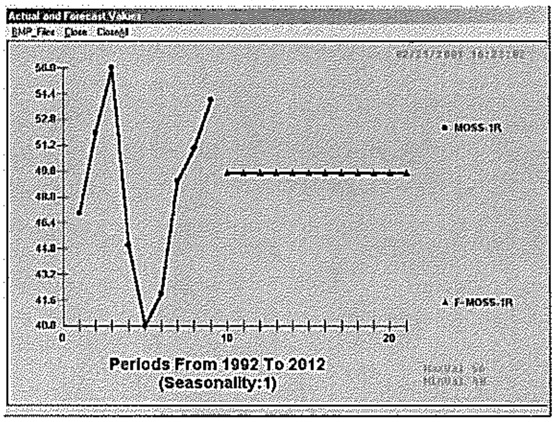
MOSS01T



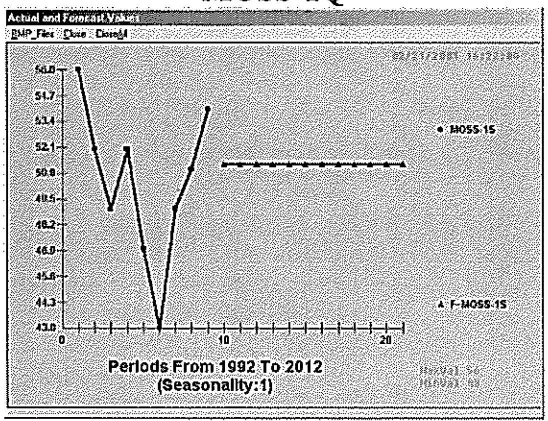
MOSS-1P



MOSS-1Q



MOSS-1R



MOSS-1S

measure the transverse profile of a pavement as a vehicle travels over it at highway speeds. Figure 2.8 is an example of the ARRB TR multilaser profilometer.

Figure 2.8 ARRB TR Multilaser Profilometer.

Each profilometer has its own unique configuration for the positioning of the elevation sensors. Figure 2.9 shows the positioning for the ARRB TR multilaser profilometer where the sensors are positioned at different spacings. By comparison, the ROMDAS TPL has 30 sensors at 100-mm equal spacings.

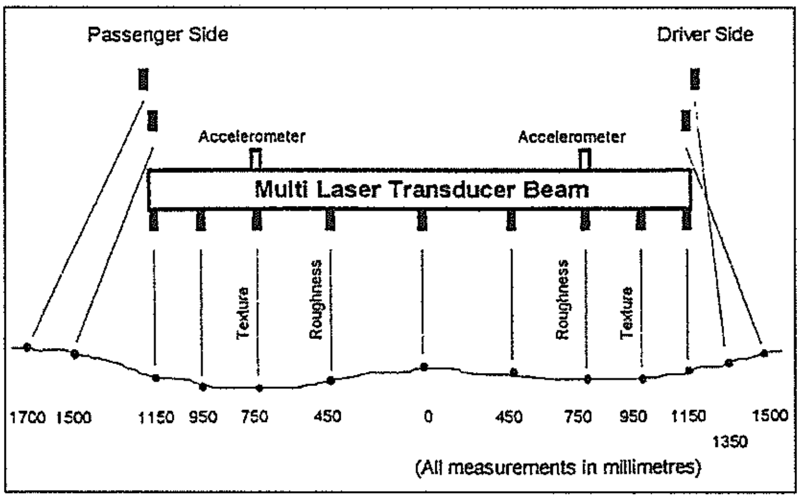


Figure 2.9 Positioning for ARRB TR Multilaser Profilometer lasers.

Irrespective of the technology used and the sensor spacing, the analytical approach is the same for all technologies. The elevations of each sensor result in the transverse profile being established, and the data are analysed to determine the rut depths.

Two basic models are used to do this: the **wire model** and the **straight-edge** model. The wire model is favoured since it is very fast in performing its calculations. It consists of stringing a simulated wire over the profile and then determine the rutting from the distance of this wire to the lowest elevation. A straight-edge model is more computationally demanding since it entails testing all the possibilities on the profile of placing the straight-edge.

measure the transverse profile of a pavement as a vehicle travels over it at highway speeds. Figure 2.8 is an example of the ARRB TR multilaser profilometer.

Figure 2.8 ARRB TR Multilaser Profilometer.

Each profilometer has its own unique configuration for the positioning of the elevation sensors. Figure 2.9 shows the positioning for the ARRB TR multilaser profilometer where the sensors are positioned at different spacings. By comparison, the ROMDAS TPL has 30 sensors at 100-mm equal spacings.

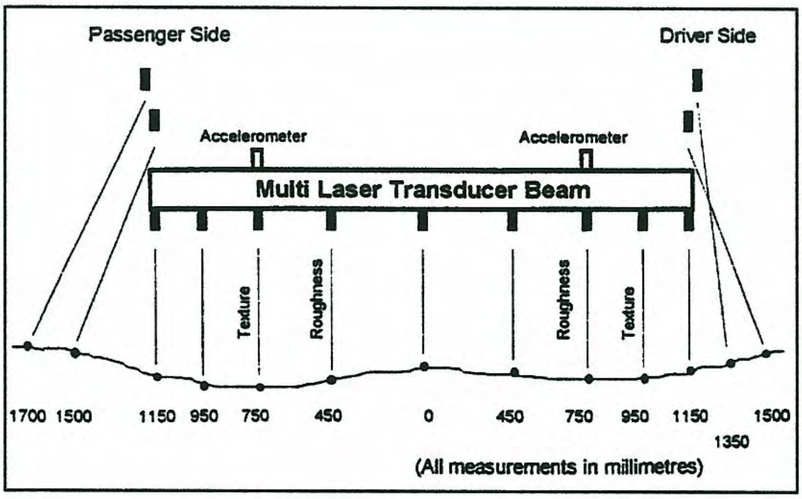


Figure 2.9 Positioning for ARRB TR Multilaser Profilometer lasers.

Irrespective of the technology used and the sensor spacing, the analytical approach is the same for all technologies. The elevations of each sensor result in the transverse profile being established, and the data are analysed to determine the rut depths.

Two basic models are used to do this: the **wire model** and the **straight-edge** model. The wire model is favoured since it is very fast in performing its calculations. It consists of stringing a simulated wire over the profile and then determine the rutting from the distance of this wire to the lowest elevation. A straight-edge model is more computationally demanding since it entails testing all the possibilities on the profile of placing the straight-edge.

Flushed surfaces have a lack of surface texture and a shiny or slick appearance. It is recorded as the length (in m) of wheelpath flushed.

Figure 2.4 Stripping of seal layer.

2.2.9 Edge Break

As shown in Figure 2.5, edge break is rated as such where the width of the sealed surface (including the sealed shoulder) is reduced by 100 mm or more from the nominal sealed edge. The length recorded shall be the length measured from the start of the taper leading up to the +100 mm edge break, to the point where the broken edge rejoins the line of the nominal seal edge.

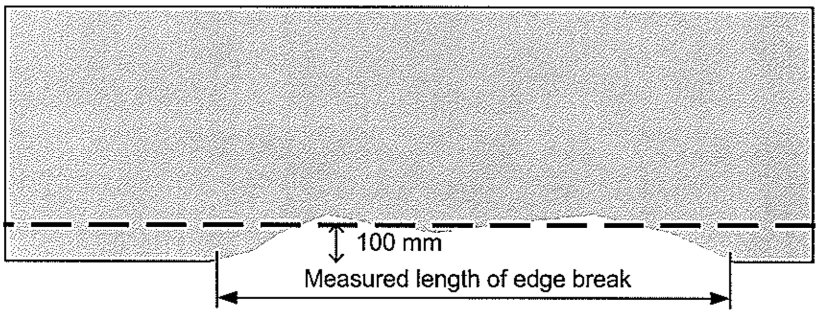


Figure 2.5 Example of edge break.

2.3 High Speed Data Collection

High speed data measurements (HSDA) are comprised of:

- Roughness,
- Rut Depth,
- Texture.

Each of these measurements is described below.

- **IFI – International Friction Index.** The IFI was the outcome of a major experiment by PIARC in 1992 to harmonise friction measurements. The IFI is composed of two numbers: the calibrated wet friction at 60 km/h (F60) and the speed constant of wet pavement friction (S_p). It is calculated from a measurement of pavement macrotexture and wet pavement friction.

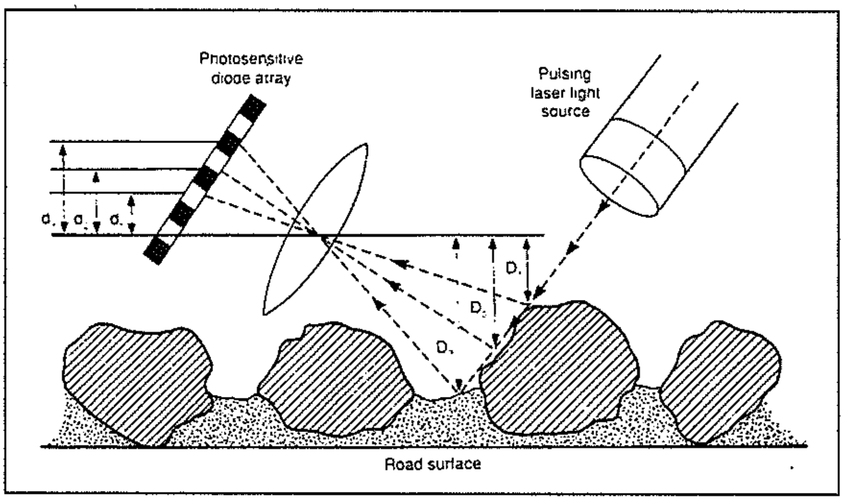


Figure 2.12 Texture measurements (Millard 1993).

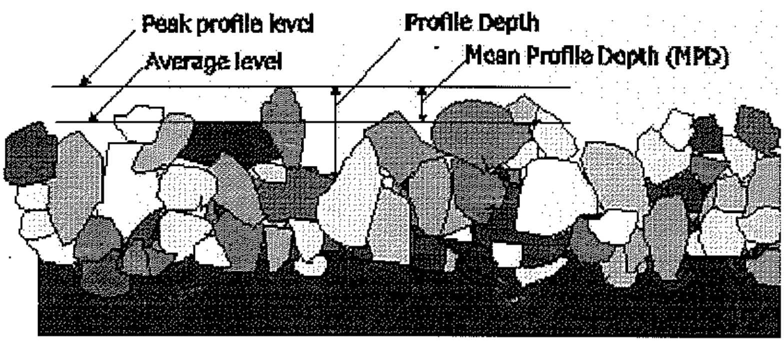


Figure 2.13 Calculating mean profile depth.

2.4 Sources of Data Errors in Road Surveys

The annual survey data represent a “time” series, where the spatial location and order of the sampling points play the role of time. The goal of the project is to develop a statistical approach, and a simple method for applying it, to evaluate these data for consistency.

There are three sources for errors in data collection:

- random errors,
- systematic errors, and/or
- operator errors.

- **Level shifts** arise when there is a shift in the magnitude of the data, for example the roughness after shape correction will be much lower than before.

Autobox develops an autoregressive model which is used to identify the unusual values. This exception analysis is suited to the objectives of this project for assessing the quality of road data since its output can be used immediately in a quality control situation. However, the small sample sizes and scatter in the data present a challenge to the analytical capabilities.

6.6 Application of Autobox Exception Analysis to Sample Roughness Data

To evaluate the capabilities of the Autobox exception analysis routine, a sample data set of 20 x 100 m sections of roads were selected from the databases: 10 for State Highways and 10 from an RCA. The data were selected so that they covered the full range of data present: little variability, high variability, large pulses (inliers and outliers), and level shifts due to maintenance. Figure 6.1 shows the data used in the analysis. It is plotted in Figures 6.2 and 6.3.

NOTE: It must be emphasised that the data used in the analysis was not typical of State Highways or local authorities. It was selected specifically because it had features which would test the capabilities of the Autobox software.

HISTORY OF SERIES

Exit

	SERIES	1992	1993	1994	1995	1996	1997	1998	1999	2000
1	*12-17-A	90	90	78	74	71	76	84	83	85
2	*12-17-B	74	74	68	80	68	63	83	82	82
3	*12-17-C	69	65	69	89	64	96	124	133	139
4	*12-17-D	96	94	94	99	110	108	121	109	112
5	*12-17-E	101	112	126	108	118	117	109	116	107
6	*12-17-F	74	80	78	113	115	121	130	133	70
7	*12-17-G	64	76	94	74	79	82	83	79	72
8	*12-17-H	94	90	108	92	95	89	98	97	66
9	*12-17-I	98	99	103	134	123	136	153	133	72
10	*12-17-J	84	90	94	117	122	133	129	141	68
11	*BOUN-5K	65	60	53	59	55	52	55	58	61
12	*BOUN-5L	52	52	63	65	39	38	70	74	77
13	*BOUN-5M	43	44	56	54	50	57	73	77	80
14	*BOUN-5N	47	44	53	50	50	49	57	60	63
15	*BOUN-5O	47	44	60	61	60	54	53	56	58
16	*MOSS01T	52	44	49	42	40	43	51	54	56
17	*MOSS-1P	61	98	71	56	57	41	46	48	51
18	*MOSS-1Q	47	60	56	48	47	42	48	50	53
19	*MOSS-1R	47	52	56	45	40	42	49	51	54
20	*MOSS-1S	56	52	49	52	47	43	49	51	54

Figure 6.1 **Sample data for testing Autobox.**

Appendix A. Precision of Roughness Measurements

In order to assess the precision of roughness measurements, data were collected from measurements in New Zealand using a response-type roughness meter, and from the published literature for profilometers.

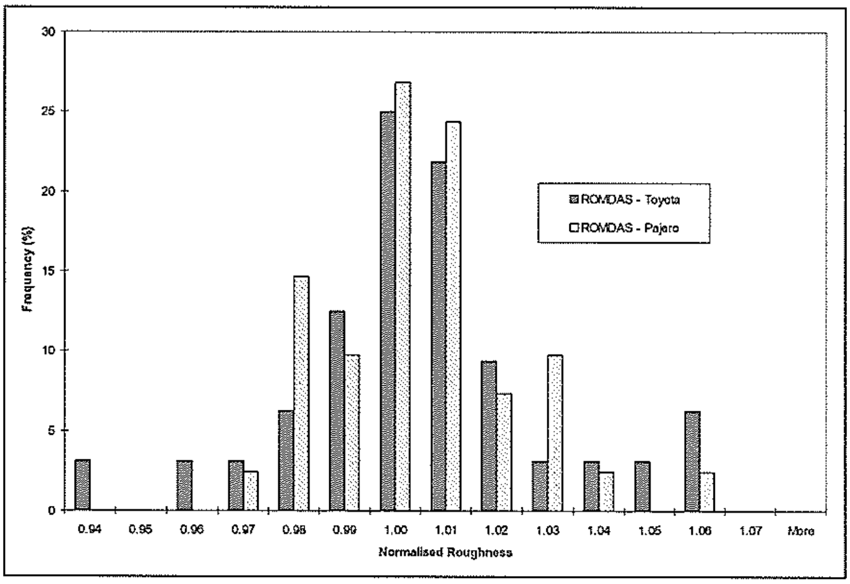
The data for the repeatability analysis came from calibrations using the two HTC ROMDAS survey vehicles: a Toyota van and a Mistubishi Pajero, hereinafter referred to as the Toyota and Pajero.

The data were collected in repeat runs over a series of test sections in August 2000 (Pajero) and November 2000 (Toyota). Different sections were used in each study, along with different drivers. Measurements were made at 50 km/h and 80 km/h.

The data consisted of the total raw roughness for both wheelpaths at each site, with a minimum of 2 runs being made at each site. A total of 41 site runs were available for the Pajero; 32 for the Toyota.

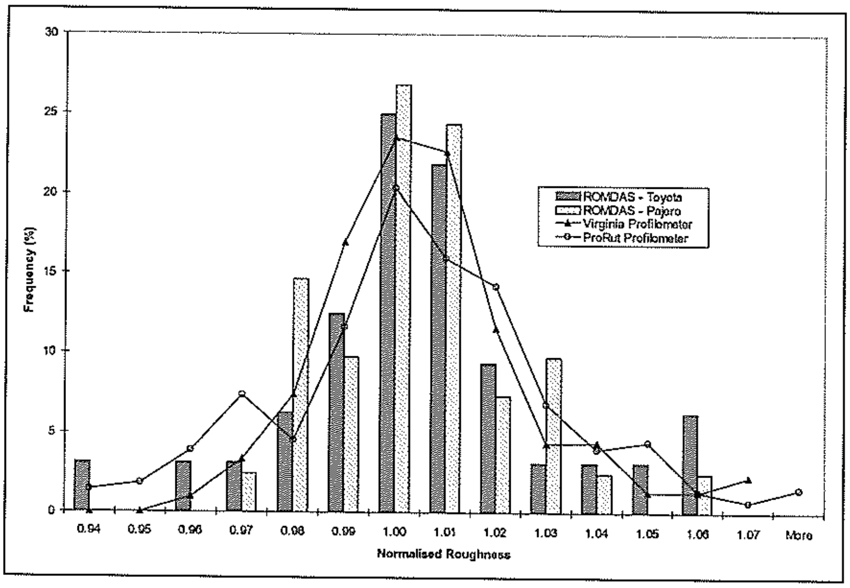
As suggested by Karamihas et al. (1999), the normalised raw roughness was adopted as the measure of repeatability. This was calculated by dividing the roughness from each run by the mean of all runs for that site. Thus, if an individual test run produced a roughness of 3000 counts/km and the average for all runs was 3050 counts/km, the normalised roughness was 0.984. The advantage of using this measure is that it is independent of the magnitude of the roughness and can therefore be used across all sites.

Figure A1 Histogram showing the distribution of the normalised roughness data.



The histogram in Figure A2 shows the ROMDAS distribution along with distributions from a ProRut profilometer (Karamihas et al. 1999) and the Virginia profilometer (McGhee 2000). With the exception of a few outliers with the Toyota, the response-type systems show equivalent repeatability to the profilometers.

Figure A2 Distributions for ROMDAS, and two profilometers.



To further investigate the precision of the ROMDAS Bump Integrator (BI), data were obtained from calibrations undertaken by ROMDAS users in different countries. For each speed and site the raw roughness was divided by the mean roughness to obtain the normalised roughness. These normalised roughnesses were then analysed to investigate the percentages within 2% and 5%. Each country represents a unique driver-vehicle combination. The results are given in Table A4.

Table A4 Results of calibrations from different countries.

Vehicle	Std Dev.	% within 2%	% within 5%
Tonga (n=48)	2.7%	56.3	93.8
Indonesia (n=144)	3.4%	56.3	90.9
Malaysia (n=50)	3.0%	64.0	88.0
Laos (n=72)	5.2%	27.8	65.3

These data show that wide variations can be expected when using the same instruments, and that these differences are a function of the driver and vehicle combination.

On the basis of the analysis here, the assumption is that response-type measurements have a standard deviation of 3% and laser profilometers of 2%.

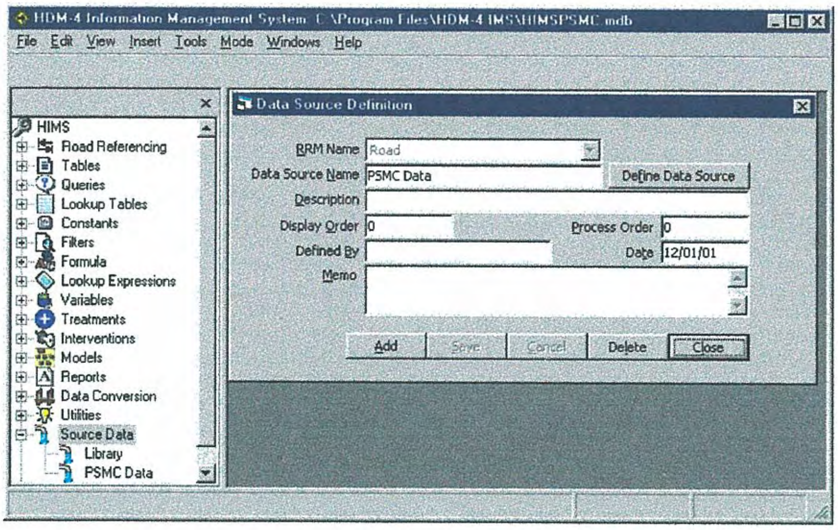


Figure B1 Defining database for source data.

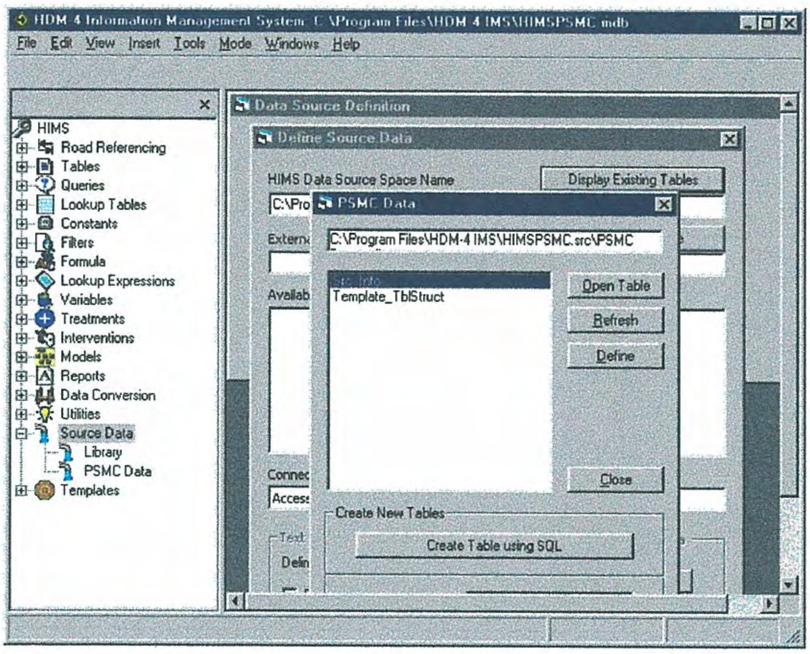


Figure B2 Display existing tables in source database.

Carriage Way (Carr_way.unl, carr_way table) → carr_way#asc
Road Name (Roadname, RoadNames table) → roadbanes#asc
Geometry (HSgeometry file, HSD_Geometry table) → geometry#asc
HSD-roughness (HSDRough.unl, HSD_Rough file) → HSD-Roughness#asc
HSD-rutting (HSDrutt.unl file, HSD_Rutting table) → HSD-Rutting#asc
HSD-Skid (Skidres.unl file, Skid_Resistance table) → HSD-Skid#asc
HSD-texture (HSDtext.unl file, HSD_Texture table) → HSD-Texture#asc
Rating (Rating.unl file, Rating table) → Rating#asc
Roughness (Rough.unl file, Rough table) → Roughness#asc
Top Surface (TopSurf.unl file, Top-Surface table) → Surface#asc
Pavement Structure (p_struct.un file, Pave_Structure table) → Pave_Structure#asc