Department of Transport and Main Roads

Guideline

Statistical cluster analysis of deflection data for pavement structural identification and design

July 2020

Queensland Government

Translating and interpreting assistance

The Queensland Government is committed to providing accessible services to Queenslanders from all cultural and linguistic backgrounds. If you have difficulty understanding this publication and need a translator, please call the Translating and Interpreting Service (TIS National) on 13 14 50 and ask them to telephone the Queensland Department of Transport and Main Roads on 13 74 68.

Feedback

Please send your feedback regarding this document to: tmr.techdocs@tmr.qld.gov.au

Contents

Figures

Equations

1 Introduction

This Guideline provides guidance on the application of statistical cluster analysis to deflection data used for pavement structural design.

Possible uses for the method include:

- assist with identification of homogenous sections within a pavement
- identify potential locations for further investigation, including cores, augers and trenches, and
- reduce the number of deflection points submitted to back-calculation processes.

Examples presented use Traffic Speed Deflectometer (TSD) data extracted from the Transport and Main Roads network-level data repository; however, the method is applicable to multivariate data from any source.

2 Background

Bulk changes to the pavement structure are recorded in as-built drawings and network-level database details such as construction date, layer thickness and material type; however, relatively minor alterations and repairs may not always make their way into such records, therefore, regular collection of network-level deflection and defect measurements is used to record more detailed structural change over time, including environmental effects.

As a rule, identical pavement structures subject to the same conditions will develop similar defects and require similar treatment; therefore, a reliable and repeatable method of identifying similar structures is required.

The data from several devices allow structural comparisons to be made:

- TSD – suited to network-level data collection,
- Falling Weight Deflectometer (FWD) – suited to project-level data collection, and
- Ground Penetrating Radar (GPR) – currently used only for project-level data collection.

Statistical cluster analysis has been applied to classification of multivariate data since the 1930s and can be used to separate data into groups from any of the listed devices.

Currently, only the FWD device permits an estimate of the layer elastic moduli required for structural design purposes and is, therefore, the preferred choice for project-level work; however, TSD data can be used to reduce the cost of FWD testing by highlighting portions of the pavement that may not require FWD testing to the same extent as other portions. Specifically, TSD data can also be used to estimate an appropriate FWD sampling interval (see Transport and Main Roads' Guideline, *Deflection Testing of Roads with Traffic Speed Devices:* request a copy of this publication by emailing ET_PMG_Director_Pavement_Rehabilitation@tmr.qld.gov.au).

Traditionally, a variety of measures were extracted from deflection data to visualise different structures within the pavement, including:

- maximum deflection (D_0)
- representative subgrade deflection (D_{900})
- curvature (D_0–D_{200}) and
- deflection ratio (D_{250}/D_0).

These measures indicate that attempts to identify structural groups within the pavement are essentially multivariate in nature.

The hierarchical agglomerative clustering used in the method includes a variety of metrics and linkage methods.

Traditionally, either individual deflection bowls or the mean bowl of a group of 'similarly shaped' deflection bowls have been used as input to back-analysis procedures. The first method relies on the assumption of very small measurement errors associated with each deflection. The second method attempts to reduce the effect of measurement errors by using the mean deflection bowl of a group. Regardless of the approach, it can be demonstrated that subsequent back-analysis will magnify such errors, typically by a factor of approximately 3.

Historically, the judgement of what constitutes a 'similarly shaped' bowl has been somewhat subjective. Cluster analysis provides a scientific (repeatable) method by which groups of bowls with similar properties can be identified.

3 Methodology

A hierarchical clustering procedure (hclust) of the R statistical computing language, available under a GNU General Public License (free), have been used to implement the clustering.

Most of the processing can be carried out using R, which is capable of interfacing to a variety of external software, should project- or company-specific formatting be required.

Bowls are compared by means of simple linear regression, which excludes the intercept. The remaining regression parameters of Slope and the Coefficient of Determination (R^2) serve as measures of the relative amplitude and shape, respectively, of the bowls being compared. This provides a dimensional reduction of n deflection measurements per bowl to just two per bowl pair.

Measurements represented on the x and y axes of the regression are drawn from the same statistical population and can be interchanged between axes. The variance of the data used for x values, therefore, has a similar variance to that used for y values. Under these circumstances, orthogonal regression is more appropriate than the Ordinary Least Squares (OLS) regression, the latter often being provided as the default method in most software. Orthogonal regression is, therefore, used to compare deflection bowls throughout this methodology.

A metric is defined as the statistical distance between observations and, since the regression parameters are defined over dissimilar ranges, Slope from $-\infty$ to $+\infty$, and R^2 from 0 to 1, they need to be transformed to provide suitable components of a metric.

Since regression variables x and y are interchangeable, the regression Slope and its reciprocal must be regarded as analogous; therefore, a convention is adopted whereby Slope values less than unity are transformed to their reciprocal value, thus providing a revised slope 9m) in the range 1 to $+\infty$.

In cluster analysis, a smaller metric implies a closer relationship between the objects being compared and a metric of 0 suggests that the objects are identical; however, the range of Slope and R^2 need to be transformed to achieve this objective.

3.1 Arithmetic transform

Although a logarithmic transform has been used in the past to convert coordinate values of 1 to 0, simple arithmetic adjustment can also be used to achieve this as follows.

An initial transformation of R^2 is indicated in Equation 3.1(a).

Equation 3.1(a) – Initial transformation of R^2

$$\boldsymbol{y = 1 - R^2}$$

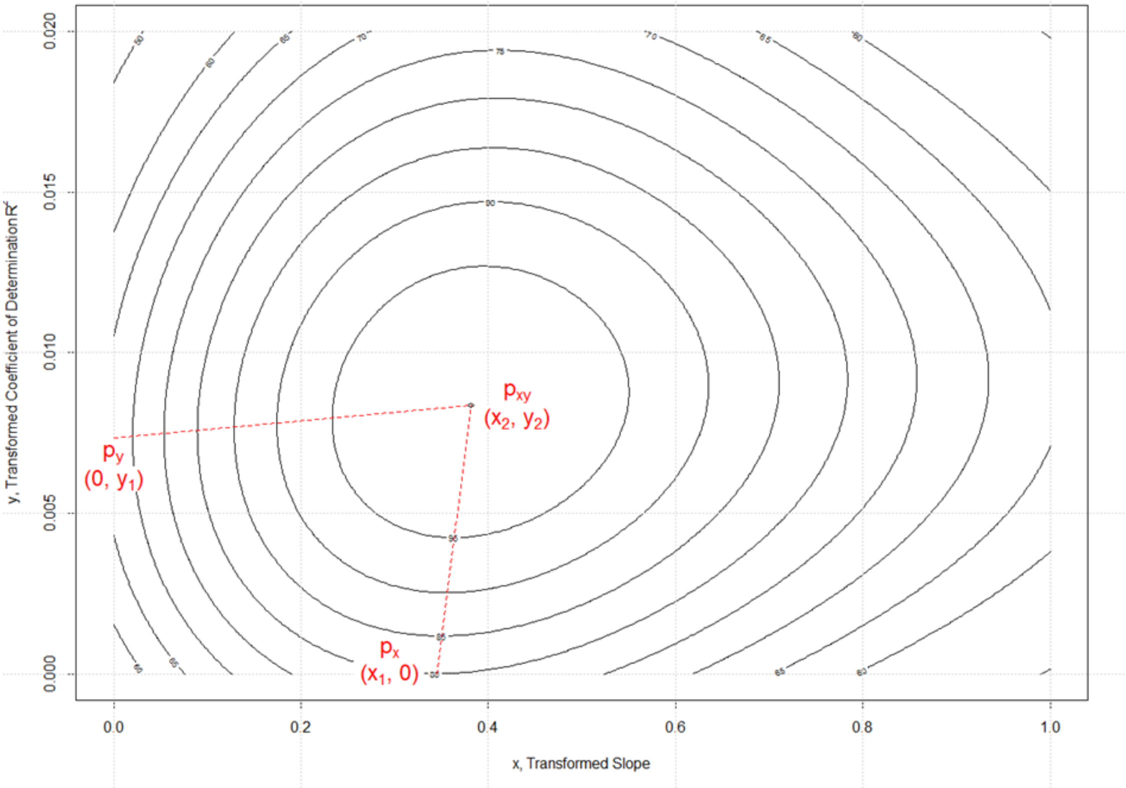
Similarly, the revised slope m is transformed as indicated in Equation 3.1(b).

Equation 3.1(b) – Initial transformation of revised slope m

$$x = m - 1$$

Using the two-dimensional kernel density estimate procedure 'kde2d' of the 'MASS' library in R, a point density object can be obtained. By plotting the point density object, using the 'contour' procedure in R, a point density contour chart similar to that of Figure 3.1 can be obtained.

Figure 3.1 – Point density contours of transformed orthogonal regression coordinates



In Figure 3.1, the maximum point density of the xy plane is indicated by point p_{xy} (x_2, y_2), and represents the most common combination (mode) of x and y coordinates found in the data. The maximum point density on the x axis is the point p_x $(x_1, 0)$, while that on the y axis is point p_y $(0, y1)$.

It is clear that combinations that lie below line $\overline{p_y\, p_{xy}}$ and to the left of line $\overline{p_x\, p_{xy}}$ lie closer to the origin (perfect match) and are more likely to provide a resilient basis for the formation of groups. The distance from the origin to these boundary lines could provide the basis for a metric, except for the fact that the cut-off threshold of a metric is typically constant, requiring a constant distance to boundaries from the origin.

By applying a further series of transformations, the orthogonal regression coordinates can be coerced to provide a constant distance from the origin to the boundaries.

3.2 Geometric transform

By multiplying x values by $1/x_2$ and y values by $1/y_2$, the coordinates of point p_{xy} become unit values (1,1) and the argument (angle from x axis) of the line $\overline{O\ p_{xy}}$ from the origin (0,0) to point p_{xy} becomes 45 degrees or 45/180×π radians.

Equation 3.2(a) can then be used to represent the revised $\overline{\mathrm{p}_y\ p_{xy}}$ line.

Equation 3.2(a) – Revised $\overline{\boldsymbol{p_y\ p_{xy}}}$ line

$$\boldsymbol{y = \frac{(y_2 - y_1) \cdot x + y_1}{y_2}}$$

Similarly, Equation 3.2(b) can be used to represent the revised $\overline{\mathrm{p}_x\ p_{xy}}$ line.

Equation 3.2(b) – Revised $\overline{\boldsymbol{p_x\ p_{xy}}}$ line

$$\boldsymbol{x = \frac{(x_2 - x_1) \cdot y + x_1}{x_2}}$$

It will be noticed that, if an x value of 1 is used in the right-hand side of Equation 3.2(a), the resulting value of y is 1. Similarly, if a y value of 1 is used in the right-hand side as in Equation 3.2(b), the resulting value of x is 1. These are the unit values described earlier.

3.3 Unit-square transform

To obtain unit-square boundaries, the line $\overline{\mathrm{p}_y\ p_{xy}}$ must be parallel to the x axis and line $\overline{\mathrm{p}_x\ p_{xy}}$ parallel to the y axis. This is achieved by dividing the revised y coordinates about the 45-degree $\overline{O\ p_{xy}}$ line by Equation 3.2(a) and the revised x coordinates to the right of the 45-degree $\overline{O\ p_{xy}}$ line by Equation 3.2(b).

This transform places a portion of the coordinates within a unit square as indicated by Figure 3.4 by the green circular symbols

3.4 Unit-circle transform

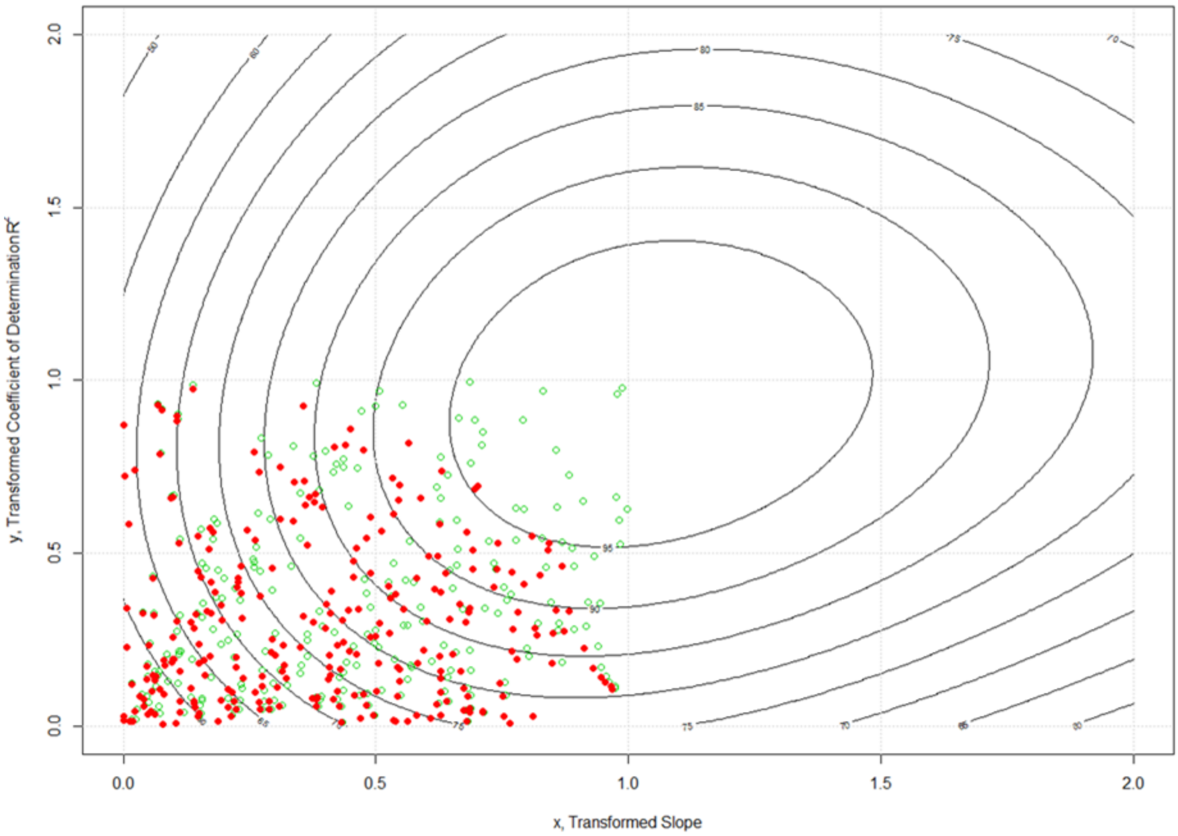
If the argument of a pair of transformed coordinates is defined as in Equation 3.4(a), and the coordinates with an argument greater than 45 degrees is multiplied by the sine of the argument and those less than 45 degrees multiplied by the cosine of the argument, the coordinates within the unit-square boundary now lie within a unit circle with the centre at the origin as indicated by the red spot symbols of Figure 3.4.

It will be noticed that coordinates close to the x and y axes remain relatively unchanged, while those with an argument close to 45 degrees are reduced by a factor of approximately $\sqrt{2}$.

Equation 3.4(a) – Argument of a pair of transformed coordinates defined

$$argument = \arctan(\frac{y}{x})$$

Figure 3.4 – Transformed coordinates



A metric ***r*** as indicated in Equation 3.4(b) can now be used with the transformed data.

Equation 3.4(b) – Metric for use with transformed data

$$r = \sqrt{x^2 + y^2} = 1$$

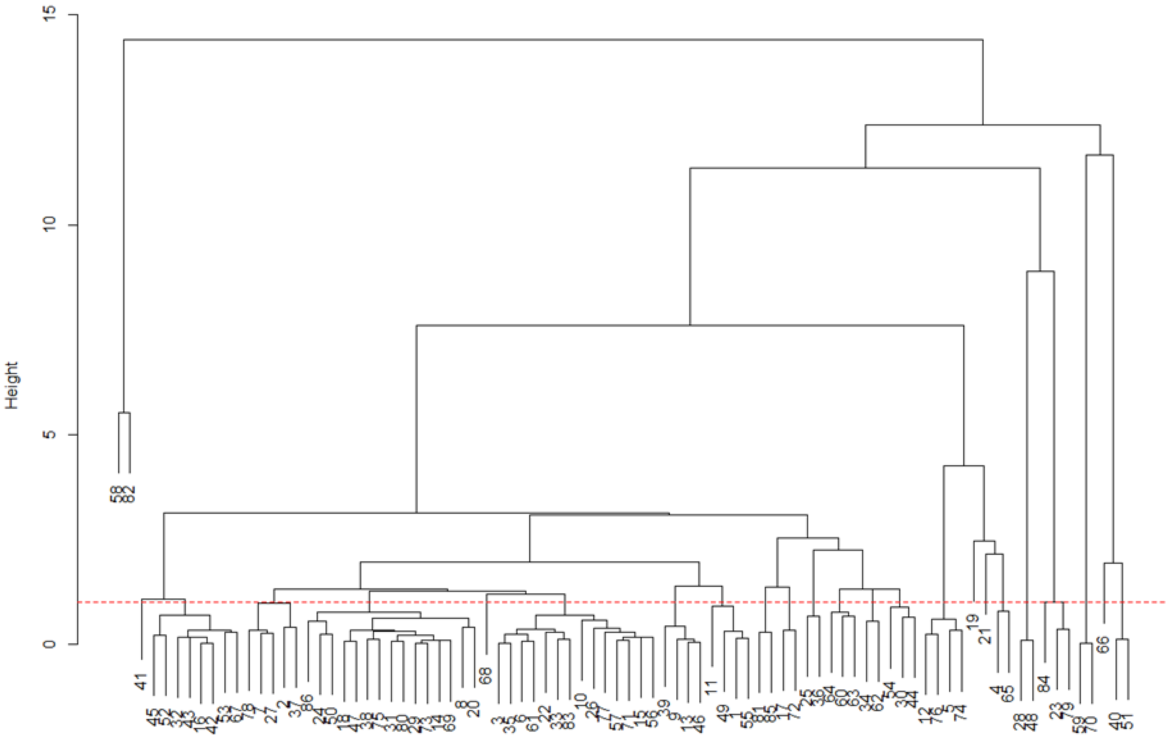
3.5 Clustering

A distance matrix composed of the metrics from all transformed coordinate combinations can now be generated as input to a hierarchical cluster algorithm.

In addition to a distance matrix, the clustering method may require a linkage method to be selected.

Figure 3.5 provides an example of the dendrogram produced by the R 'hclust' procedure with a 'centroid' linkage method. The horizontal red dashed line represents the cut-off threshold for the metric defined in Equation 3.4(b).

Figure 3.5(a) – Dendrogram



The R 'hclust' procedure offers the choice of eight linkage methods: 'average', 'centroid', 'complete', 'mcquitty', 'median', 'single', 'ward.D' and 'ward.D2'. The question then arises as to which of these methods should be used.

Figure 3.5(b) compares the number of groups produced by the R 'cutree' procedure for each linkage method using the unit cut off threshold of Equation 3.4(b) and Figure 3.5(a).

As indicated, the 'single' linkage method tends produce a small number of large groups with relatively diverse properties in each group. By way of contrast, the 'complete', 'ward.D' and 'ward.D2' methods tend to produce a large number of small groups with relatively consistent properties in each group, with the possible disadvantage that the large number of groups may be less suitable for visualisation purposes. Based on linkage method details, group constituency, and the mid-range number of groups produced, the 'centroid' method is recommended.

Figure 3.5(b) – Comparison of linkage methods

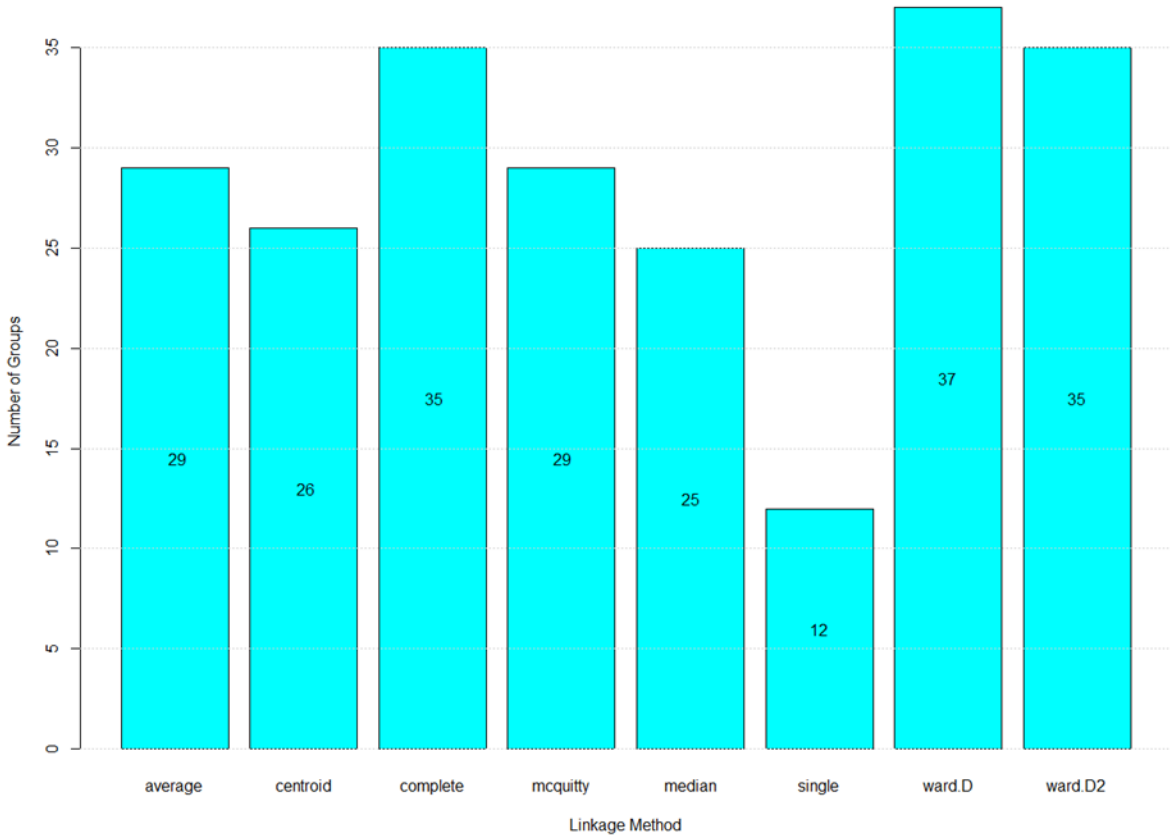


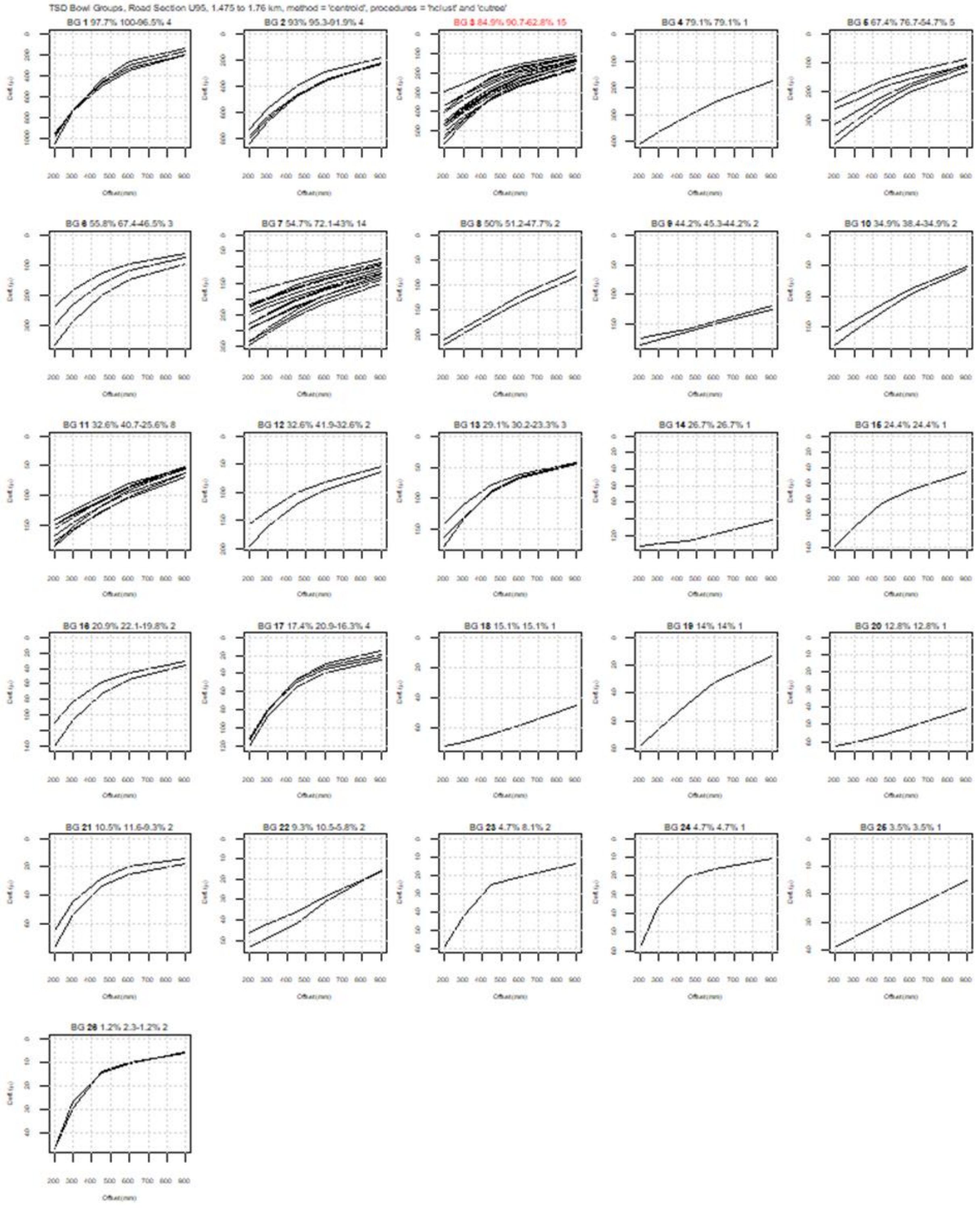
Figure3.5(c) contains an example of a 'bowl group thumbnail' report. The purpose of this report is to allow the user to assess group consistency; that is, each group should contain bowls of similar shape and magnitude. If not, the cut-off threshold can be reduced or a linkage method which favours a larger number of groups could be selected; however, the unit cut-off and 'centroid' linkage method will be suitable for most data.

The groups are numbered in ascending sequence according to the most common (mode) magnitude in the group and presented in the same order; thus, bowl group 1 will contain bowls that are generally of greater magnitude than the succeeding groups.

The header of each chart contains the group number, mode percentile, percentile range and number of bowls included in the group. Headers that contain the 90th percentile deflection, often used for pavement design purposes, are highlighted in red.

Since TSD data have been used as the basis for the examples, only measured deflections are included, estimated deflections such as D_0, D_{750}, D_{1200} and D_{1500} have been excluded. Where FWD data are used, no deflections are estimated, allowing a full range of deflections to be included in the analysis.

Figure 3.5(c) – Bowl group thumbnails



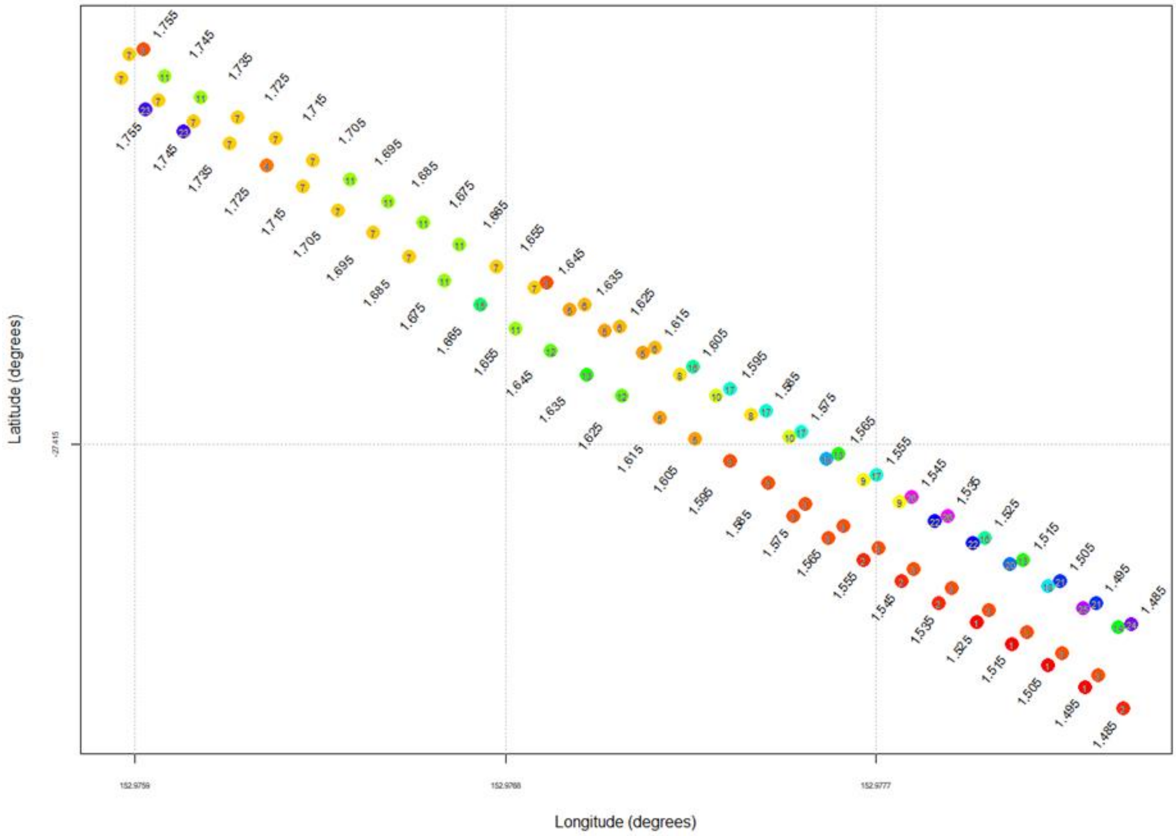
From Figure 3.5(c), bowl groups 3, 7 and 11 are relatively large and will influence how the pavement is subsequently sectioned and treated, while smaller bowl groups may be investigated for possible localised repairs prior to more extensive treatments.

Note: Lower-deflection groups often contain a greater variety of shapes than higher-deflection groups. This may relate to defects such as fatigue cracking which are often indicated by a sudden change in the magnitude of deflection at a particular offset from the load, implying poor Load Transfer Efficiency (LTE).

Figure 3.5(d) indicates how bowl groups may be viewed in a geospatial context. To cater for users with colour perception issues, both colour and labels have been used to differentiate between bowl groups. The red end of the spectrum has been used to indicate higher-deflection groups while the blue / magenta end of the spectrum indicates lower-deflection bowl groups.

Note: A label placed within a symbol should use the [complementary colour] of the symbol to highlight the label.

Figure 3.5(d) – Geospatial view of bowl groups



4 Back analysis procedure

Currently available back calculation software was developed to estimate layer elastic moduli from FWD data. Back calculation of TSD data should not be attempted with this software unless the data has been converted to an FWD equivalent (see Transport and Main Roads' Guideline, *Deflection Testing of Roads with Traffic Speed Devices:* request a copy of this publication by emailing ET_PMG_Director_Pavement_Rehabilitation@tmr.qld.gov.au).

The following steps should be used to back calculate grouped FWD data.

1. Capture deflection data to a format suitable for processing. Where data have been collected using several plate loads (for example, 40, 60 and 80 kN), all loads can also be used to separate bowls into groups, effectively adding load sensitivity to R^2 and slope regression parameters; for example, an FWD device equipped with nine geophones and tested at three plate loads can provide 27 regression points with which to compare deflection bowls.

2. Separate test points into sites where a structural difference is known or likely. Sites are typically identified based on a gap in chainage (longitudinal), but can also be defined based on carriageway, lane or even wheelpath (transverse) in conjunction with statistical cluster analysis.

3. Identify wheelpaths which require a Deflection Moisture Adjustment Factor (DMAF) (see Transport and Main Roads *Pavement Rehabilitation Manual*) to be applied prior to back analysis.
4. Separate each site into longitudinal sections based on relative deflection levels, historical records and GPR results. To assist with sectioning, deflection data may be further processed to produce a Cumulative Difference plot. It is acceptable to use several wheelpaths in the sectioning process.
5. Commence statistical clustering of bowl shapes in each section using the methods described in Section 3.
6. Select bowl groups for back analysis. Groups containing a single bowl may include significant geophone measurement errors; therefore, groups containing at least two bowls (preferably more) are recommended. The bowl group containing the 90th percentile deflection is of particular interest as the results will likely influence the design of the pavement. Both inner and outer wheelpaths should be sampled to allow the effects of moisture adjustment to be represented. Large bowl groups with lower deflection should be examined to assess the feasibility of design options proposed for higher deflection groups and vice versa.
7. Select a representative bowl from each group. The bowl with the most common D_0 deflection (mode) should be used.
8. Equation 4 (see *Pavement Rehabilitation Manual* Figure 2.13.10.4.6) can be used to assess the validity of the back calculated subgrade modulus from a FWD device with a 40 kN load, and representative subgrade deflection D_{900r}. According to Austroads' *Guide to Pavement Technology Part 2*, Equation 2, the subgrade modulus should be approximately ten times the subgrade California Bearing Ratio (CBR).

Equation 4 – Validity of back calculated subgrade modulus

$$CBR = \frac{0.5911}{{D_{900_r}}^{1.4759}}$$

Note: Most pavements exhibiting approximate linear elastic behaviour allowing deflections for loads other than 40 kN can be 'normalised' to 40 kN by multiplying by 40 and dividing by the measured load. The range of validity for this approximation is listed in the *Pavement Rehabilitation Manual*.

9. The moisture adjustment method of the *Pavement Rehabilitation Manual* should be applied to wheelpaths, except that the method is confined to bowl groups rather than sections; therefore, the bowl group or groups selected for back analysis should include both inner and outer wheelpaths. Each selected bowl group can, thus, be represented by a moisture adjusted representative bowl.
10. Submit the moisture adjusted representative bowl of each selected group to a back analysis process.
11. Bowl groups can also be used in a geospatial context to assist in the identification of locations for cores, augers and trenches (refer Figure 3.5(d)).