Data confidence for effective highway asset management

Greater understanding of pavement condition data is being built with methods developed to improve data quality and analysis. This paper is by Matthew Byrne, Steve Biczysko and Tony Parry.

Introduction
In recent years there has been increased focus on the implementation and subsequent financial benefits of highway asset management. To maximise the benefits of measurements of pavement condition, it is imperative that the quality of this data is understood, not only for the analysis of current data but further to improve the quality of future collected data. This work on data quality fits into the overall Nottingham Transportation Engineering Centre asset management research, which seeks to improve the overall quality of data analysis and prediction.

Benefits of increased data quality
Research has traditionally focused on developing more accurate models of pavement condition deterioration. The following discussion does not attempt to detract from the importance of these developments, but rather seeks to highlight that typical deterioration modelling only solves half of the puzzle of predicting how a pavement deteriorates over time.

For a deterioration model, assuming a simple linear case, the required two points are the ‘now’ (or present) condition and the ‘then’ (future) condition. Reducing the error in the ‘then’ prediction reduces the value α; the error associated with the prediction of condition change (Figure 1a).

Research in this area is highly valuable since economic benefits from even small reductions in α can be very significant.

Most deterioration models require the ‘now’ value or starting point. If this value is incorrect, see error β (Figure 1b) which will in turn increase the total error of prediction, α + β. This combined error may be further increased as some deterioration models include ‘now’ as a parameter in estimating the change in condition. In this case the final error of prediction will not be α + β, but rather α x Ω[β + β] where Ω describes the relative contribution β makes towards predicted change in condition. Effective highway asset management needs to rely on minimising not only the error range α but the complete error range α + β.

Identifying current condition and error estimates
Minimising the error range β can be achieved by correctly analysing the historical data to identify the likely true current condition, ie, identifying and removing measurement errors. The apparent solution to achieving the best inference of current condition is to analyse the complete historical dataset simultaneously to identify how it changes with respect to distance and time. This approach has been successfully developed at the University of Nottingham and is referred to as the Minimum Message Length Two Dimensional Segments (MML 2DS).

Tests and simulated trials have shown this outperforms alternative techniques to correctly infer true progression rates.

The MML 2DS process not only improves the estimate of true current condition, but further contains an advanced system to identify the error of each measurement. It does this to remove bias on estimates of condition, but has the further benefit of providing output which can be analysed separately to extract highly valuable information.

Visual identification of outlier trends
The MML 2DS calculates the probability by which each measurement is an outlier or error. By visualising this with a colour sequence abnormalities in data can be identified. If the measurement is below that expected by the MML 2DS trends, it is coded as blue (with the strength of the colour increasing with the probability of it being an outlier) and vice versa if the measurement is above that expected with a red colour.
This colour coding is shown in Figure 2. The resulting output of the complete dataset error is displayed in Figure 3. Errors are to be expected but they should be randomly distributed. Any trends in errors (which are visualised as adjacent bands of similar colour) highlight those sectors for closer causal examination.

Figure 3 highlights some groupings of possible errors which will influence the data quality. Group 1 (around 220-500 metres in 2005-1) has a consistent red coding. This defines this area of data as measuring much higher than expected given the adjacent years’ values. This apparent collection error is visible in Figure 4 which shows the actual measured RHS rutting for years 2003 to 2005 with the 2005-1 measurements significantly higher than the previous two years, which are closely aligned.

Group 2 differs from Group 1 as it is aligned along the time and not distance axis. Figure 4 shows these values and they appear to be not measurement error but rather a small section (each year has this error) which is performing far worse than the adjacent sections. In this case this is not identifying data error but rather what appears an unusually poor performing section of the total road. Group 3 describes the year 2004-1 as appearing to underestimate the likely condition level given the 2003-1 and 2005-1 measurements.

Simultaneous analysis of the complete dataset across distance and time is the only way that these errors could be identified. Further, it highlights the risk of estimating current condition from only the previous year’s data measurements.

**Conclusions**

Correctly identifying the true condition not only provides the important means of reducing the error in the overall predicted future condition, but also provides the basis for identification of likely measurement errors. Once any trends in errors are identified the means to reduce their future occurrences is available. Measured data is critical to good asset management but any errors in this data must be identified and if possible removed from the analysis for cost effective highway engineering solutions to be developed.

**Xtra-info: www.**