Mining Road Condition Data using Minimum Message Length

Rawle Prince

Nottingham Transportation Engineering Centre (NTEC),
Division of Infrastructure & Geomatics
Faculty of Engineering
University of Nottingham

KTP Associate, Devon County Council

April 25, 2013

Outline

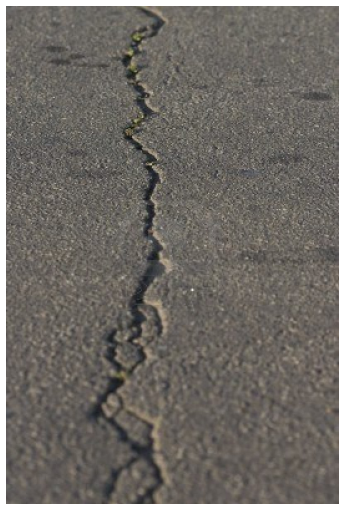
- Overview of Road Condition Data Collection
- Very Brief Overview of Minimum Message Length (MML)
- Addressing Limitations of Original Prototype System
- Summary

Road Condition Data Collection

Road agencies collect expensive data forming the backbone of the asset management system to identify various indicators, for example:

- Rutting

- Cracking:



Road Condition Data Collection

This data collected is often subject to noise, errors and other issues such as:

- unrecorded maintenance
- changes in the measurement devices, and
- possible seasonal variation.

All these combine to make identifying current condition and true progression rates difficult.

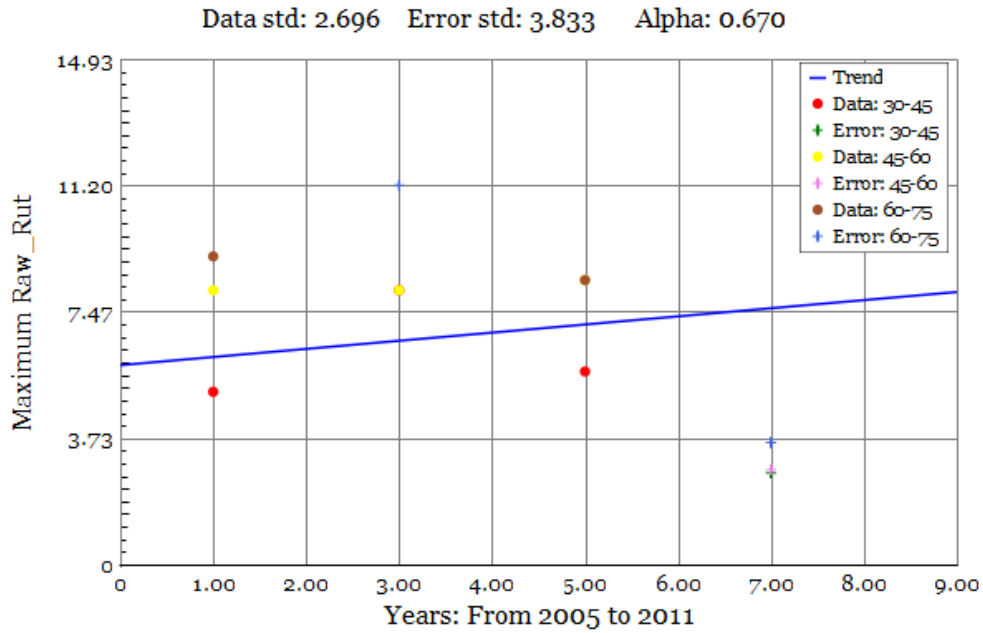
Sections and chains

A road network is typically subdivided into a series of links/sections, each of which is subdivided into standard-lengths subsections/chains.

- Data from sections and/chains are analysed to identify various indicators of performance and maintenance needs,
- and to calculate progression rates and pavement deterioration models.

Analysis Issues

- Progression rates often disregard trends with negative slopes:
  - a change in condition should be explained by a maintenance intervention and/or outliers identified as errors



- Progression rates are used to seed deterioration models:
  - Small variations in the starting value can significantly impact the outcome; error identification is crucial
- NTEC has developed a technique to identify progression rates, maintenance interventions and errors using the Minimum Message Length (MML) metric

The MML Metric

MML is a model comparison technique, residing at the intersection of Information Theory and Statistics

- Employs the metaphor of sending information through a communication channel
- In the case of (linear) regression, can be thought of as least squares plus a few *bells* and *whistles*

The *bells* and *whistles*

- Typically, we may use a least squares estimate to fit trends to data. In MML this *cost* is augmented by others inversely proportional to:
  - **Bells:** the believed goodness-of-fit of the model — i.e. the cost of encoding the model structure
  - **Whistles:** the model complexity — i.e. the cost of encoding the model parameters

The MML Metric

These costs are combined to form the *message length* of a given model, given the data:

- The *sender* sends the bells and whistles, stating her belief about a given model
- The *receiver* responds with the least squared estimate, indicating how good the sent model actually is

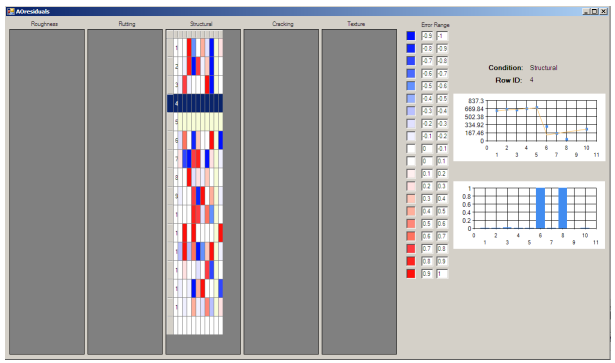
So, the goodness-of-fit of a complex model (usually in terms of the number of parameters) is penalized by costs relating to its complexity.

- Hence, MML is deemed to endorse Occam's Razor: even when models are not equal in accuracy to the observed data, the one generating the shortest overall message is the one most likely to be correct

Analysing Road Condition Data using MML

The original prototype was developed by Matthew Byrne, a colleague at NTEC:

- Included a data sharing technique to compensate for limited data – i.e. data from adjacent chains are combined to from *maintenance groups*.
- Used a "DNA diagram" to display errors



Limitations

The original prototype was very slow

- took weeks to terminate on a section with twenty five chains
- very inconsistent – did not give the same result on repeated runs

I was tasked with addressing the limitations and implementing a system to analyse DCC's road condition data

Optimisations

The optimisations done on the algorithm used in the prototype system include:

- initialis the linear parameters with estimates based on the linear regression MML function (think the least squared function without errors/weights)
- replace the constraint solver for negative slopes with a 'best' estimate based in reversing functions with negative slopes
- employ 'smart'techniques for identifying maintenance groups, rather than evaluating all possibilities
- exploit various programming techniques, e.g. parallelism

Results

- fast and consistent – repeated runs give the same results!

Figure : Preliminary results on small sets of DDC's Rutting data

Summary

- Analysing road condition data is important for the calculation of progression rates and for deterioration modeling.
- The approach used at NTEC is to employ MML to identify progression rates, maintenance interventions and errors.
- Significant improvements were made to the original prototype and we are currently finalizing, as far as we are aware, the first commercial application of MML for DCC to anlayse their road condition data

Thank You