**Queensland Rail (QR) Data**

**Background Information**

TRC = track recording car, data relates to status of the rail e.g. METRAGE = position on the track measurement was taken, TOP L/Top R = position of the left and right rail, TWIST refers to twist in the track (e.g. left rail is higher than right etc)

TRC operates quarterly and surveys the entire QR network in Queensland. Changes in the position of the rail and twist are used by the engineers to inform where maintenance work needs to be undertaken.

A primary driver for rail movement is ballast. This project would focus on identifying the need to “undercut” i.e. dig under and replace ballast. This process is expensive, time-consuming and disruptive.

GPR = geo-survey undertaken every ~3 years. Identifies PVC = Percentage Void Contamination and measures contamination in the ballast (e.g. mud holes, plant growth) and ballast density.

**Current/Historic Practice**

At present QR undertakes a manual visual review of the TRC data and compares this to the GPR data to see whether there are “hot spots” requiring maintenance work.

One initiative attempted by the engineers is to produce a “heatmap” in Excel. This was undertaken for a small (e.g. ~9km) section of the Shorncliffe line (one direction up or down).

The heatmap “HEATMAPS 20m.xls” was produced as a prototype and it attempts to identify degradation in TRC condition and present GPR data for the same section of track.

Given the number of datapoints e.g. every 5m for ~9,000m of track and many standard deviation calculations, the spreadsheet is relatively unstable and difficult to work with. It is not feasible to use Excel for larger sections of the rail network.

**Automation of Heatmaps**

Given the time QR has provided me and to gain a better understanding of the data, I have attempted to provide a Python process to automate the process of producing Heatmaps: “QRHeatMap.py”.

This can produce heatmaps for ~9km section in several minutes and as the output spreadsheets do not contain calculations, they are stable.

**DVA Team Actions and Opportunities**

QR are currently reviewing the automated heatmaps to determine what additional insight they would seek if they had this information for the entire network.

I have asked about the potential for introducing a ML model to utilise the TRC and GPR data to make predictions.

The objective of the model would be to predict asset condition (as measured by top of line) at a future time (e.g. 12-24 months).

As such, we have requested the following additional information which were identified as a set of variables that could potentially be used to train a model:

1. Culvert & embankments (proxy for drainage)
2. Level crossings
3. Versign L&R (available in TRC data)
4. GPR along line (1, +/-5, +/-10) (available in GPR data)
5. Train load (#trains, tonnage per month)
6. Maintenance – tamping (e.g. $ p.a amortised)
7. Maintenance – undercutting (e.g. $ p.a amortised)
8. 3 years of TRC – we could possibly train using data from 2 years ago then test based on most recent TRC data i.e. the model will not have seen the most recent runs so is a good assessment as to its predictive power (if any)
9. Speed changes
10. Changes in modulus of track (bridges/level crossings) – track hard points (bridges/level crossings)
11. Track structure
12. Weather (precipitation and/or extremes in temperature) – probably low priority given it is unlikely to relate to sections of track and is not within control, but could be considered

Items in green are almost certainly available in the timeframe for the project.

Items in orange should be/are likely to be available.

Items in black are unknown – of these I believe item 1 would be useful: culverts correlate with drainage issues and poor drainage is a major cause of ballast issues. As such, I expect culvert information will assist in the predictive power of the model.

**Proposed Model Options**

Suggestions as to how we may build a ML model:

1. Attempt to train a model using the above parameters from data over past 36 months
2. Model could be trained:
3. Over time i.e. comparing different time periods in the TRC, and/or
4. Along the track e.g. does a degradation in track condition impact the quality of line further down the track (e.g. train bounces)
5. Test predictive ability based on the most recent TRC data
6. If the model demonstrates it has any predictive power, predict track condition in future (e.g. 2 years’ time) given various scenarios for the predictors, such as investment in maintenance etc.
7. Introduce visualisation e.g. geographic drill-down t obtain specific track condition

**Next Steps**

1. QR to advise on what data is available – expect by 4th October.
2. DVA team to confirm it wishes to proceed with the project