PREDICTIVE MAINTENANCE APPROACHES BASED ON CONTINUOUS MONITORING SYSTEMS AT RIO TINTO

Cameron Thompson¹, Paul Reichl¹, Dingyang Zeng¹, Joshua White¹, Faez Ahmed², Hitanshu Sethi³

¹Institute of Railway Technology, Monash University, AUSTRALIA
²University of Maryland, USA (Previously at Rio Tinto Ltd, AUSTRALIA)
³Rio Tinto Ltd, AUSTRALIA

Corresponding Author: Cameron.Thompson@monash.edu

SUMMARY

Irregular track geometry can incite undesirable vehicle dynamic response modes that increase track loading, reduce component life and increase the risk of vehicle derailment. Geometric irregularities in track can typically be identified by monitoring wagon-track dynamic activity. Instrumented ore car (IOC) continuous monitoring systems measure a range of response metrics including spring nest deflection, which under nominal service loads and speeds, are a key indicator of geometry induced wagon-track dynamic activity.

This study demonstrates the benefits of predictive maintenance approaches facilitated by the trending of continuously measured performance data, as developed for Rio Tinto’s heavy haul iron ore railway network in Australia’s Pilbara region. Predictive maintenance approaches facilitate the improvement of maintenance planning operations resulting in better track surface and line condition, reduced risk to infrastructure and rolling-stock as well a reduction in the need for reactive temporary speed restrictions (TSRs) and unscheduled maintenance activities.

1. INTRODUCTION

In order to maintain nominal track geometry throughout a railway system, it is important to perform regular track maintenance. Maintaining correct track geometry condition is an important aspect of railway operation because irregular track geometry can incite undesirable vehicle dynamic response modes that increase track loading, reduce infrastructure component life and increase the risk of vehicle derailment.

Settlement of ballast formations, caused by the cyclic loading and vibration from network traffic, can exacerbate the degradation of track geometry and should therefore be systematically maintained.

Track geometry issues commonly trigger a wide range of track maintenance operations and renewals. Tamping is an effective maintenance procedure which repacks ballast particles under sleepers in order to restore the correct geometrical position of ballasted tracks and is often used when geometrical issues are first identified. Historically speaking, maintenance strategies have been subjectively based on the knowledge and experience of infrastructure operators. However, the continuous drive for increased efficiencies within the heavy haul industry necessitates the need for a greater understanding of maintenance performance and intelligent, data driven approaches that can target and prioritise maintenance activities.

1.1 Track Geometry

Track geometry degradation caused by the settlement of ballast is a complex phenomenon affected by many factors including duration of operational exposure, dynamic loads, usage intensity and climatic events [1, 2, and 3].

Track geometry irregularities can be classified into different groups depending on their wavelength and type of irregularity: isolated (single point) defects, short wavelength irregularities (1-5m), and long wavelength irregularities (>5m). Generally speaking, long wavelength irregularities are more problematic at high speeds, where they can induce swaying and bounce mode motion of the wagon body.

Tamping of ballast is a common form of remedial action that is performed to repair or correct track geometry irregularities caused by the settlement of
ballast with the aim being to bring the track geometry back to nominal levels.

1.2 Tamping Maintenance

Tamping can be explained as the compaction of the ballast to increase its supportive effect around and under the sleepers. The tamping and lining machine has tamping tynes that are inserted into the ballast on either side of the sleeper. Lifting rollers then raise each sleeper to the target level which also creates a space under the sleeper. The tamping tools then squeeze the ballast to adjust the position of individual particles in order to fill any cavities. Simultaneously, the lining tool of the tamping machine will adjust the position of both rails so that the track is straightened. A typical tamping sequence is shown in Figure 1.

![Figure 1: Sequence of Tamping [4]](image)

The quality of track geometry can be assessed in several ways, including: manual assessment during maintenance work, visual assessment via track inspectors or train drivers or with an automated track recording car.

Geometric irregularities in track can also be identified by monitoring wagon-track dynamic activity using instrumented ore car (IOC) continuous monitoring systems.

1.3 IOC Continuous Monitoring Systems

IOC systems are fully automated measurement platforms that are embedded on standard ore wagon revenue vehicles. These vehicles are permanently equipped with advanced measuring systems including different types of sensors and logging units to provide continuous feedback on both rail condition and train operation. The primary use of IOCs to date has been for track condition monitoring. A Rio Tinto IOC is shown in Figure 2.

![Figure 2: Instrumented ore car (IOC)](image)

Rio Tinto has a number of IOCs which are in operation throughout its heavy haul iron ore railway network in the Pilbara region in the north of Western Australia.

Some key advantages of an IOC over existing track geometry measurement vehicles include:

- IOCs reduce the need for track downtime as they measure the condition of the system as part of normal railway operations.
- The IOC measures dynamic responses under the same axle loads as nominal fleet vehicles.
- IOCs are cheaper to purchase and operate.
- IOCs provide a more frequent and continuous coverage of the rail network and near real time feedback of track and train issues.
- Multiple units can be used on a train consist and across the network.
- Additional measuring equipment can be easily added to the flexible platform.

The ability of the IOC to directly communicate over mobile and satellite networks greatly increases the detection to response time for critical maintenance activities. IOC field recordings are automatically downloaded to the main logging unit where the data is then remotely transmitted to the data processing centre at the Institute of Railway Technology, Monash University in Melbourne for further analysis and reporting.

2. PREDICTIVE MAINTENANCE APPROACH

Continuous monitoring systems have enabled a paradigm shift in maintenance planning for the railway industry by facilitating the shift from reactive, to predictive maintenance.

Reactive maintenance can be characterised as the often costly, unscheduled or emergency correction of track conditions to nominal levels. Predictive maintenance on the other hand, is pre-scheduled maintenance activities that are prioritised to locations exhibiting rapid deterioration rates or
where track availability needs to be scheduled well in advance.

Data driven, predictive maintenance approaches enable improved planning and scheduling of maintenance and minimises the risk to infrastructure and rolling-stock, whilst also reducing the need for temporary speed restrictions (TSRs).

2.1 Measured Parameters

In order to optimise maintenance strategies through the use of predictive approaches, there needs to be a consistent method to measure track degradation rates and the consequences of the degradation in terms of cost and risk. Obtaining timely and accurate knowledge about degradation rates helps to estimate the right time for inspections, maintenance and renewals. The measured data can also be used to identify critical parts of a railway network for targeted maintenance intervention.

Data collected from Rio Tinto’s numerous IOC continuous monitoring systems was used to facilitate the predictive maintenance approaches demonstrated in this paper.

IOCs are typically equipped to measure:

- Bogie spring nest deflections to detect track geometry irregularities.
- Unsprung accelerations to detect track surface defects.
- Lateral accelerations for train instability.
- Coupler forces and brake pipe pressures for in-train forces and driving strategies.
- GPS position with sub-metre accuracy for identifying the location on track.

As track geometry issues tend to be associated with longer wavelength dynamic activity, spring nest displacement (SND) measurements are likely to be particularly informative. IOCs are capable of continuously measuring SND, otherwise described as the vertical travelling distance of suspension springs (mm), under nominal service loads and speeds, which is a key indicator of track geometry induced wagon-track dynamic activity. It is also therefore possible to make use of historical data at fixed locations on track to estimate how the responses are changing at these locations over time. Indeed it is the ability to trend the responses with time that enables an approximate track geometry degradation rate to be determined.

IOCs are configured to capture SND at each of the spring nests as illustrated in Figure 3. This sensor arrangement is used as it allows for the bounce (pitch mode) and the body rock (roll mode) of the wagon motion to be measured at each end of the wagon.

In the current study, the recorded SND dynamic response and speed data is mapped against track chainage (km) using the differential GPS position for each of the IOC trips in reference to detailed track lookup tables. The entire track network was then subdivided into discrete track bins/sections spanning approximately 50m, where each bin maintains a unique track identification number or UID. Rio Tinto’s railway network totals approximately 40,000 bins.

For the purposes of the predictive maintenance approaches demonstrated in this paper, only the combined maximum SND response was used for each discrete track section. The SND response value can therefore be defined as:

\[
SND_{BIN} = \text{Max}_{\text{a}}(SND_1, SND_2, SND_3, SND_4) \quad (1)
\]

where $SND_a$ denotes the spring nest travel distance (mm) at spring nest corner number ‘n’ as illustrated in Figure 3. Similarly, only the maximum speed was used for each discrete 50m track bin.

\[
\text{Speed}_{BIN} = \text{Max}(\text{Speed}) \quad (2)
\]

2.2 Database

The vast amounts of data collected lends itself well to a database solution. IOC trips are automatically downloaded and post-processed. The processed data and metadata, spanning several years, is then stored in a database. This opens up significant possibilities as it facilitates a level of
consistency and acts as an enabler for a host statistical analysis and data mining.

2.3 Trending Track Degradation

The continuous monitoring capabilities of the IOC system ensures that the dataset is always current and provides a rolling time history of SND dynamic response values that can subsequently be used to determine the localised rate of track geometry degradation, with respect to time, for each discrete track bin. In the current study, the localised rate of track degradation was determined using a least squares linear regression trend (line of best fit) through the time history data for each track bin.

An example linear regression trend is shown in Figure 4, where the SND response (mm, top) and speed (km/h, bottom) are plotted against the date. This linear regression trend line allows for the response data to be extrapolated out to a future date in order to assess the likely response magnitude at that time. If the projected magnitude for the location exceeds a predetermined response threshold level, with a sufficiently high level of confidence, then the location can subsequently be flagged as one that requires some form of intervention.

As a result, it would generally be expected that any such maintenance activities would result in a step change in response profiles with respect to time. An example of this type of behaviour is illustrated in Figure 5, which exhibits the classical 'saw tooth' shaped response profile in line with known tamping dates (red vertical bars).

![Figure 5: Post-tamping linear regression trend](image)

It is therefore clear that any predictive models would need to take into account previous maintenance activities so that only the deterioration rate since the last maintenance intervention is considered. To account for this, tamping maintenance records were incorporated into the regression calculations so that predictions were made based only on data since the last known tamping activity, or where no previously recorded tamping activities in the preceding 6 months.

One of the obvious extensions of including tamping records is that it is also possible to estimate tamping effectiveness based on the magnitudes of the responses before and after tamping.

2.5 Regression Models

With the prior maintenance information included, it is then possible to calculate the track degradation rate for each of the approximate 40,000 50m bins spanning the entire track network.

Each regression calculation has a number of coefficients that describe the fit. In the case of the simple linear curve fitting model these coefficients include metrics such as the slope, the intercept, the \( R^2 \) value, or 'goodness-of-fit', along with the

---

Figure 4: IOC Spring Nest Deflection (mm) and Speed (km/h) vs. time for discrete track bins
number of data points (ndp) used in the regression calculation.

For the single variable linear regression model used in the current study, that only accounts for changes in responses with respect to time (i.e. the impact of other factors such as speed, weather, etc. are ignored), it may be necessary to filter some of the predictions to better account for variations in the underlying data. In such instances the choice of the filters applied is likely to have some bearing upon how well these models predict what is actually happening.

To assure the reliability and accuracy of the predictions, only track segments with a sufficient number of data points and sufficient regression confidence were considered. For example, only those locations predicted to reach the predefined threshold within the specified forecast period that have a $R^2$ value greater than 0.5 might be considered. Or similarly, only those locations showing rapid deterioration (slope) and where a measured response has already been observed to exceed some fraction of the allowable thresholds might be considered.

Other common considerations include known IOC operational differences. For example, due to the fact that a wagon’s dynamic response to a track irregularity is more pronounced when the wagon is loaded, it may be desirable to treat both the empty and loaded trips separately and/or together.

Another operational consideration is to separate out the various wagon classes in order to account for their relative response sensitivity to track geometry inputs. This was found to be an important step as each class has its own unique suspension and damping characteristics. A class that may have a low dynamic sensitivity to a given track geometry input can therefore obscure high response trends from other more sensitive wagon classes. Categorizing or separating out the classes may further help to identify significant trends at a location.

One of the advantages of filtering the responses based on a set of criteria is that it also allows each of the locations flagged to be assigned a level of urgency based on either the model coefficients or other additional extremal measures, which can then be used to further prioritise any planned maintenance activities.

The forecast response thresholds were set to some fraction of a response that may constitute a reasonable derailment risk to rolling stock or damage to infrastructure. This response threshold is commonly referred to as a Severity 1 threshold.

In the current study the actual threshold limits chosen were based on existing experience in operating the IOCs systems on Rio Tinto’s railway network. For illustrative purposes the SND response is commonly normalised against the Severity 1 threshold so that a value of 1.0 corresponds to a Severity 1 magnitude response.

$$SND_{NORM} = \left( \frac{SND}{\text{Severity 1 Threshold}} \right)$$ (3)

Figure 7 shows post-tamping linear regression trend lines for loaded IOC trips separated by wagon class. The lighter shade data points indicate the empty IOC trips that have very low comparative response magnitudes. This graph also includes reference vertical lines (grey) showing the date of viewing the data (today) and the forecast date.

For all these cases the question then becomes how to appropriately determine the conditions on which to filter the predicted responses so as to only highlight those locations that are deteriorating? In addition, data should not include responses that appear to be due to underlying variability in the curve fit. To try and determine the best way in which to filter and then rank the predictions based on their deterioration rate, retrospective testing was used.

![Figure 7: Loaded wagon linear regression trends separated by wagon class.](image-url)
3. RETROSPECTIVE TESTING

As with many things, the first step in trying to optimize the performance of a system (be that track, rolling stock or a combination of both) is measuring how it performs against a set of benchmarks and constraints and then monitoring how this changes in response to input changes.

In this instance the forecasting or predictive accuracy was determined using retrospective testing. Retrospective testing makes use of historical data, so that forecasts can be made for track locations at a series of historical dates and the accuracy of these forecasts can then be compared with actual data that was collected during the subsequent forecast period.

3.1 Hit Rate and Optimization

To quantify the impact of any changes made to the forecasting algorithms, a “hit rate” metric, or key performance indicator (KPI) was established. This KPI compared the number of locations that were forecast to reach or exceed the predefined threshold response to those locations where an exceedance actually occurred.

Retrospective testing was then conducted at fortnightly intervals throughout the historical dataset with a fixed forecast interval in order to optimise the overall hit rate. This process was also subject to the constraint of attempting to minimize the number of incorrect predictions (i.e. those locations where we predict intervention as being required but where nothing was subsequently observed). As this was essentially an optimization problem, an optimization approach was used to determine the coefficients that should be used to get the best hit rate subject to the applied constraints.

4. APPLICATION AND PERFORMANCE

4.1 Surface and Line Prediction (SLP) Reports

Using the optimised modelling parameters, it was then possible to generate regular Surface and Line Prediction reports. These reports highlighted and ranked locations on track where the trends in the measured IOC data suggested that a response exceeding the predefined threshold could be observed within the next forecast period.

The ranking of the locations was achieved by considering the degradation rate, the number of data points used in obtaining the trend, the aggregate recent maximum response and other external factors such as any recent track geometry exceptions in the near vicinity.

4.2 Prediction Accuracy

Continued development of the models has seen the prediction accuracy improve over time, with the number of locations being flagged for inspection simultaneously dropping.

The plot in Figure 8 shows the monthly percentage of the threshold exceedances that occurred that were predicted at some point in the 12 weeks prior to the exceedance event occurring. Figure 8 also shows the average number of locations that were flagged each fortnight. These results are fairly significant as they suggest that in November, some form of warning was given for 73% of the threshold exceedances that month.

4.3 Integration with other Systems

While all of the prediction information is clearly valuable, what helps make it of even more use is that it can be integrated with other systems. This enables work orders and other remedial activities to be planned. In this case, the flagged locations were formatted in such a way that they could be directly uploaded as defects into Rail Asset Management software such as Ramsys.
5. IMPROVED MODELS AND MULTI-VARIABLE REGRESSION

Some of the key drawbacks associated with using single variable simple linear regression, even when used in combination with a range of other logic decision trees, is that the speed dependence of the measured dynamic responses is not accounted for and the uncertainty in the response is assumed to stay the same regardless of the size of the response itself (i.e. the responses are assumed to be homoscedastic).

To better handle this and other factors, such as the dependence on weather, it often makes sense to look at models that are designed to treat such dependencies. In these instances, GLM-type models or ones determined from machine learning algorithms may prove more useful. It should be noted that in the current study only linear models with additional logic based constraints were used.

6. CONCLUSION

Continuous monitoring systems have enabled a paradigm shift in maintenance planning for the railway industry by facilitating the shift from reactive, to predictive maintenance.

By approximating track degradation rates, this data-driven predictive maintenance approach has been used to identify locations on track that require maintenance intervention. This enables improved planning and scheduling of maintenance, minimises the risk to infrastructure and rolling-stock and simultaneously reduces the need for temporary speed restrictions.

Current industry application in the form of surface and line prediction reporting for Rio Tinto’s heavy haul iron ore railway network has yielded fairly significant results and continued development of algorithms has seen the prediction accuracy continue to improve over time.

7. REFERENCES