Graphical methods for railway track condition assessment and prognostics

Bjarne Bergquist (bjarne.bergquist@ltu.se)
Quality Technology, Luleå University of Technology, Luleå Sweden

Peter Söderholm
Swedish Transport Administration (Trafikverket), Luleå, Sweden, and
Quality Technology, Luleå University of Technology, Luleå Sweden

Abstract
This paper presents graphical methods for monitoring, diagnostics, and prognostics of the condition of railway infrastructure as support to maintenance planning. The paper also uses graphics to aid univariate, bivariate and multivariate analyses of large datasets of secondary data for linear asset condition assessment in the temporal and spatial domains. We present graphical methods useful for evaluating how the asset degrades and how maintenance actions affect the track condition within different time horizons. Hence, the infrastructure manager and the contracted entrepreneur can share a common view of the asset's current and future condition, as well as maintenance effectiveness.

Keywords: Graphical methods, Prognostics, Railway infrastructure

Introduction
In Sweden and elsewhere, condition assessments are increasingly often used for decisions on when and where to maintain different systems of the railway infrastructure, e.g. power supply, signalling, track, and catenary systems. Special measurement trains that regularly measure important variables are currently the predominant way to perform inspections to obtain condition measurement data for the track and the catenary systems. The infrastructure manager or contracted entrepreneurs plan for the needed maintenance when measures surpass certain threshold limits. The importance of the variable and the severity of the deviation from the designed limit dictates the time bounds before actions are required, as well as the type of action itself. In Sweden, such times can range from acute, requiring immediate corrective traffic management and maintenance actions, up to planned maintenance actions within three years after detection (Stenström, 2015). Thus, measurements trigger much of the maintenance and its planning. If the timetable (i.e., set 18 months before execution) does not allow inclusion of the required maintenance, there is a significant risk that the required maintenance action disturbs the railway traffic, requiring line shutdowns, train delays or cancellations. Condition measurements are thus of utmost importance. However, current diagnostic practice only uses measurements from the last passage of the train to monitor the asset's current condition.

However, to use only the last measurement makes the maintenance decision more susceptible to measurement uncertainty. Uncertain measurements do on the one hand force the infrastructure manager to lower maintenance threshold limits to maintain low system risks for accidents due to undetected faults. On the other hand, this practice increases the false alarm rates.

Additionally, many maintenance actions may themselves reduce the life of the infrastructure items (Arasteh Khouy, 2013; Famurewa, 2013), so the maintenance cost is not limited to the replacement costs of items worn by use and of costs for personnel and machinery (Patra et al., 2009). One example is ballast tamping; an action performed to realign the track. Tamping involves lifting the rails and sleepers while pushing a fork like equipment into the ballast beside the sleepers. The forks will then vibrate, which will fluidise the ballast so that it will flow in under the sleepers. When the machine lowers the track, tamped segments will be higher, which will hopefully reduce some geometry faults. Tamping is expensive, and the vibrating forks crush some of the ballast rocks, reducing ballast life and at the same time producing fine material, which reduces the dewatering properties of the superstructure. A rule of thumb used by the Swedish Transport Administration is that the ballast can withstand ten tampings before it needs replacement or ballast cleaning; maintenance actions that are classified as reinvestments and should improve the condition in such a way that it is considered to restart the life cycle of the ballast.

In the north of Sweden, geometrical deviations may indeed be real, but temporary. A temporary displacement of the track may be due to ground frost, which in turn, is a condition that is due to water in the superstructure or substructure. Hence, a temporary speed restriction until conditions settle may be the proper action for some geometrical deviations, rather than to perform maintenance and possibly worsen the situation. Underestimating (false alarms) or overestimating (undetected faults) will both degrade the condition decision quality thus increasing cost and degrade system usability. (Bergquist & Söderholm, 2012, 2014)

Better data analysis and better asset condition knowledge thus become essential. When the maintenance personnel acquire asset condition data, they usually screen them for causes for alarm, that is, to check whether some assets need attention shortly. Those areas where there is an acute need, or where it is evident that things are about to become acute are further studied, and those areas will be scheduled for maintenance. The rest of the measurement data will likely be stored away for later reference. However, an approach that stops with storing data for later without anyone analysing them is only a waste of storage space. A better way is to use data to help gain an understanding of the asset condition and how the condition changes over time. Data analytics in general (Levrat et al., 2008; Karim et al. 2009) and the reoccurring measurements allow for time series estimations of how the track conditions evolve (Bergquist and Söderholm, 2016). Additionally, it is possible to display, extract and analyse vital information from large datasets. Hence, condition assessment based on proper methodologies and technologies could substantially reduce maintenance-related costs and increase infrastructure capacity while maintaining traffic safety (Arasteh Khouy et al., 2014; Soleimanmeigouni et al., 2018).

Aim

This paper aims to demonstrate how graphic representations and modelling may aid monitoring, data quality assessment, diagnostics and prognostics of the track condition to support asset management maintenance planning.

Method

The method we use is to study stored condition monitoring data for asset measurements, and here we study railway track geometry data.

The studied case, the Boden – Luleå track section

The data for this study stems from measurements of the Swedish Iron ore line. Measurement cars measure the geometry with some regularity and with current regulations, the measurements are performed six times yearly. Four different measurement trains and trollies produced the data between 2007 and 2018. The measurement trains and trollies have somewhat different measurement performance and measurement profiles, such as top speed of measurements, and they have different axle loads.

The studied track section 119 links the cities of Boden and Luleå and is 35 km in length. Besides the Luleå and Boden stations, the section also has stations at a regional hospital in Sunderbyn, a freight terminal station in Gammelstad and a commuter station at Notviken. Track 119 is relatively straight and does not include short radius curves. The track is a single track with meeting point sidetracks, and both passenger and goods traffic use it. Track 119 is a heavy haul line classified to handle a maximum of 32.5 metric tonnes axle load.

Dataset issues

The data analysis of historical data usually starts with an analysis of the data itself; that is, a study of the data quality. One way of discussing big data is the 4V-model of Katal et al. (2013), comprising of data volume, velocity, variety and value. To those, a fifth V have been added, veracity (e.g. Lukoianova & Rubon, 2014). The value of the data depends on their trustworthiness. The dataset in this paper relates to numerical data, in which data veracity may refer to erroneous measurements, and the objectivity of the data and the collection systematics are not problematic per se. However, the veracity may be hurt from poorly calibrated or faulty instruments, erroneous data pre-processing and erroneous data positioning when several data sources are combined to one dataset.

Data may thus be erroneous, may be missing, and measurement data always contain noise. In all likelihood, the data needs cleaning before the data quality is sufficient for the analysis task. Often, the quality will remain too low, for instance, because of a too low information-to-noise ratio even after the analyst has cleaned the data. The risk that the database does not contain all that is relevant to the goals of the study is always present. Assume that the analyst will use the data for a regression task to find correlations among properties or variables. If the database lacks information about important events, the regression analysis will, at best, be poor or fail to find significant correlations. At worst, the correlations will be strong and significant, but pointing in the wrong directions. Data correlations may just be reactions to an external, but unknown signal, so rather than each other's causes, both are reactions to the underlying event.

The analyst extracting data from numeric databases should be aware that the data often is secondary, i.e. it was collected for other purposes than the intended goals of the present analysis. Using secondary data means that the analyst must scrutinise the original objectives to see if they interfere with the questions that the analyst is trying to answer. If the original goal may have been to supply measurements for a highly critical process where accuracy is paramount, and where one would like to spot any errors immediately, great trust in the data may be warranted. Even so, also safety-critical data contain errors that are readily obvious when they are scrutinised from a new vantage point. The analyst must also check the appropriateness of the data from other viewpoints, such as if the statistical properties of the data are compatible with the necessary analyses techniques that are to be employed.

Poor accuracy may also result from organisational gaps, for instance, because the user of the data has not made the data supplier aware that there are issues with them. Such issues may, for instance, not being forwarded to the correct recipients since the information channels are not working correctly. Proper use of the statistical methods relies on the assumption that data are independent and normally distributed. Data seldom are, but one can often salvage datasets where these assumptions are violated by choosing proper statistical tools or data transformations.

Data quality assurance in the studied case

The data for the current graphical approach to railway track condition monitoring entails many of the previously discussed issues. They are secondary data since they were sampled to evaluate the current asset status, rather than to form the basis for condition prognostics, that is, for forecasting purposes. The ultimate purpose of the analysis is to forecast asset conditions with enough accuracy so that appropriate maintenance planning can commence far enough into the future so that train schedules are not disturbed.

While this distinction between assessing asset condition and obtaining condition data useful for forecasts may seem like a minor one, there are issues with the data that a diagnostic analysis of the current asset status will not reveal. Diagnostics of the current state assessment is a snapshot of the asset condition, but it is not a motion picture, whereby one could estimate how the condition will develop over time. Prognostics require both. The measurement trains measure with differing times between measurements, and the measurements need to be linked together to evaluate the speed of change of the measured property. A reasonable way to do this that will allow for is to estimate conditions within regular intervals, based on some method, for instance by using interpolation of historical data, for instance using splines, linear regression, or just to use the latest measurements as the best estimated for the condition at a specific time. All methods have their strengths and weaknesses. Here we used averaging of measurements obtained within the latest quarter to estimate the condition, due to simplicity and extrapolation stability.

To link measurements over time means that one needs to certify that the second record relates to the same asset or asset segment as the first record. Better still if measurements have been obtained with the same equipment and personnel, with updated calibrations. To acknowledge such errors, one needs a frame of reference, that a single measurement will not provide. It is, therefore, not surprising that condition datasets obtained for a particular asset within a short time difference can differ substantially due to measurement issues rather than condition variation. Maintenance personnel did not find such issues when they only used the last and most current measurements.

Positioning is, due to the linking of measurements in time, a central property. Since we are discussing railway assets that often remain in position for tens of years (for instance track) and sometimes hundreds of years (for instance, bridges and tunnels), one may assume that positioning is trivial. If the spatial information of the data is correct, it probably easy to link measurements. If not, the analyst needs to use some strategies for overcoming the spatial errors to estimate the speed of change of the properties. In the studied case, positioning errors have been considerable, and the strategy to overcome them has been to use data binning to split the track into segments long enough for failures to have a high probability of falling into the same segment during consecutive measurements, see also Bergquist & Söderholm (2012, 2015).

Another potential difficulty is that some external disturbances have affected the asset or the measurement equipment between measurements, which would lead to errors in the rate of asset degradation. Usually, maintenance prognostics aim to predict when the condition of the asset has degraded to such a poor state that it needs maintenance. The degradation rate of the asset property is, therefore, often what one seeks to combine this information with the state itself for condition predictions. However, sometimes the recorded asset condition has improved between measurements. Some of these unexpected events may be due to maintenance actions that one can find in other databases. If there is a viable reason for an improved asset condition, the proper way to handle things could be to remove prognoses until the deterioration rate is reasonably stable. A procedure to monitor improvements could include triggering alarms for larger improvements than a threshold level based on the expected normal variation. Improvements that are larger than what can be expected just by measurement uncertainties and which cannot be explained by external factors would mean that the current model is not valid for describing current data. In the case of the subarctic railway track maintenance studied here, such improvements that are known to influence readings could be ground frost effects, maintenance or due to changes of the measurement equipment such as calibrations. Some detective work may reveal the actual causes. Generally, events that are due to pure measurement errors, such as zero recordings, are easy to find and remove.

Other unusual observations may also be easy to determine as erroneous, such as when the recordings are orders of magnitude different from the regular measurements. A multivariable bi-variate scatterplot as depicted in Figure 1 is useful to get an overview of the data. Two of the studied variables, two standard deviations of twist with 3 m or 6 m base has had several recordings stemming from another distribution, which is an illustration of the data with different orders of magnitude. The twist fault that can lead to derailments and it is thus safety critical and regularly measured. Figure 2 shows this deviation in detail. The twist data also contains zeros, which given the segmentation approach and the standard deviation studied here used is almost impossible, and the analyst should remove such observations. The time interval where ground frost events are likely should be possible to estimate for railways using local knowledge.

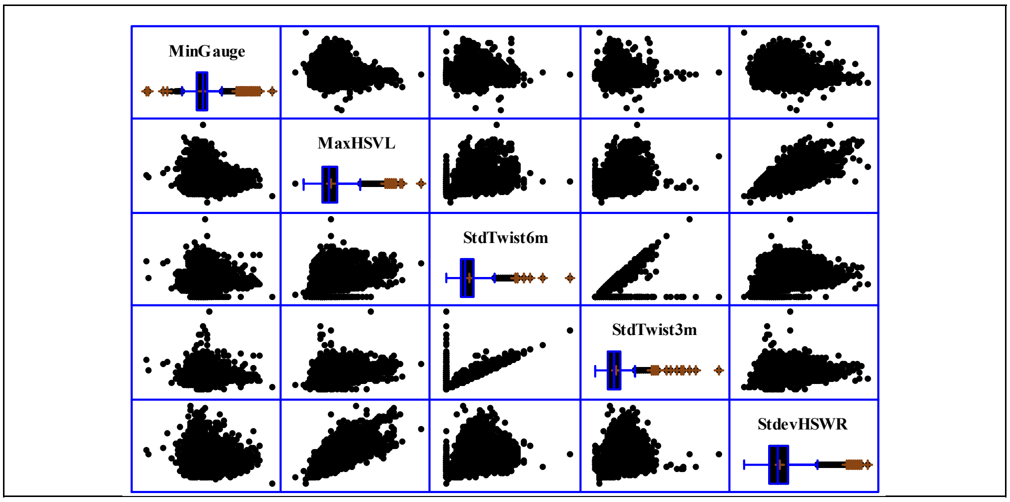

Figure 1. Matrix plot of some measured track properties important for railway track safety, equipment reliability and passenger comfort. Note that the twist measurements split correlations into two groups.

In many cases, strange observations will remain mysterious. The analyst should consider all possibilities and consequences for keeping or disposing of the strange data where the culprits are not obvious. If, for instance, the intended use of the data is condition prognostics, a restart of the prognostic model could help, depending on the model design.

For track geometry conditions, the maintenance itself will often unsettle the track superstructure. The first or first few measurements after maintenance may indicate that the superstructure geometry is well within its targets, but then the conditions may degrade rapidly until the superstructure has obtained a new, predictable, deterioration rate. If this happens, it may be advisable to restart the model. If the empirical prognostic model uses recent data to estimate future conditions of the property, the model is likely to generate poor predictions until the asset has settled. In the track case, a steady condition deterioration indicates that that the track has been subjected to sufficient transported weight or temperature variations for the ballast to wiggle into a more stable configuration.

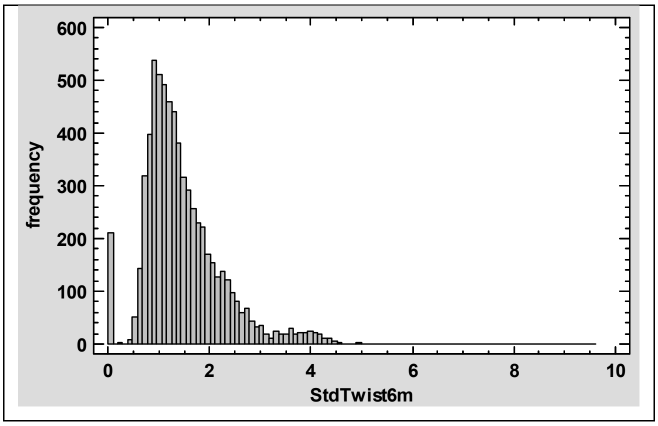

Figure 2. Histogram of the 6 m twist standard deviation reveals that two distributions have been combined, one centred just above zero. Further studies show that part of the data has a mean that is orders of magnitude lower than the rest, which indicates erroneous measurements.

Seasonal effects may be necessary to include in the prognoses models. As already stated, reoccurring events such as winter frost may influence conditions. Geometrical conditions after the spring thaw or the autumn frost seasons may improve by themselves until the next measurement since frost is a function of the local water content of the substructure and temperature. Frost may, therefore, affect the track substructure or superstructure differently with considerable local variation; differences that dissipate as the frost thaws or has settled over the whole track segment. Temperatures themselves are also important as they force the material to expand and contract, causing stresses in the linear assets due to different thermal expansion coefficients.

All of the above effects can be studied using time series plots. Figure 3 shows a time series plot of one of the critical geometrical properties, twist over 3m. The figure displays both the measurement data (circles) and quarter means (crosses). The figure also shows a Kalman filter model including a prediction confidence interval for the model based on the variation in the sample, as well as an upper confidence interval of the model data. The Kalman filter is based on the quarterly average since the sampling is not regular. The measurement frequency is too low and too irregular, so the model does not allow a standard seasonality component of the time series, but the Kalman filter is used to assess the current state but also predict future conditions. The model does in this case a poor job of estimating the data behaviour. The vertical dashed lines represent known maintenance (tamping). The twist variable must not exceed certain threshold values for comfort and safety, so large absolute values are problematic. The expectancy is that maintenance will reduce the absolute twist, at least in the short term. Studying the observations in this plot reveals that in 2009, the absolute value decreased without obvious reasons. Since the model needs to have some degree of robustness against random variation such as measurement noise, it did not respond to the rapid decline, nor did it pick up the rapid degradation rate between 2010 and 2015. The 2009 improvement suggests some maintenance action that was not recorded in the data. Interestingly, the 2013 tamping does not appear to have improved the twist. A speculation is that the degradation rate has lessened as a result of the tamping. The 2014 tampings did, however, reduce the absolute value of the twist. Judging from the data, it appears that the situation again improved in 2016, so it is possible that had been maintained or that there was something done with the measurement equipment, but this is not clear from available data. Another possibility is that the track measurement systems were updated or calibrated during these episodes. The situation improved again in 2017, to such a large extent that the model had automatically restarted.

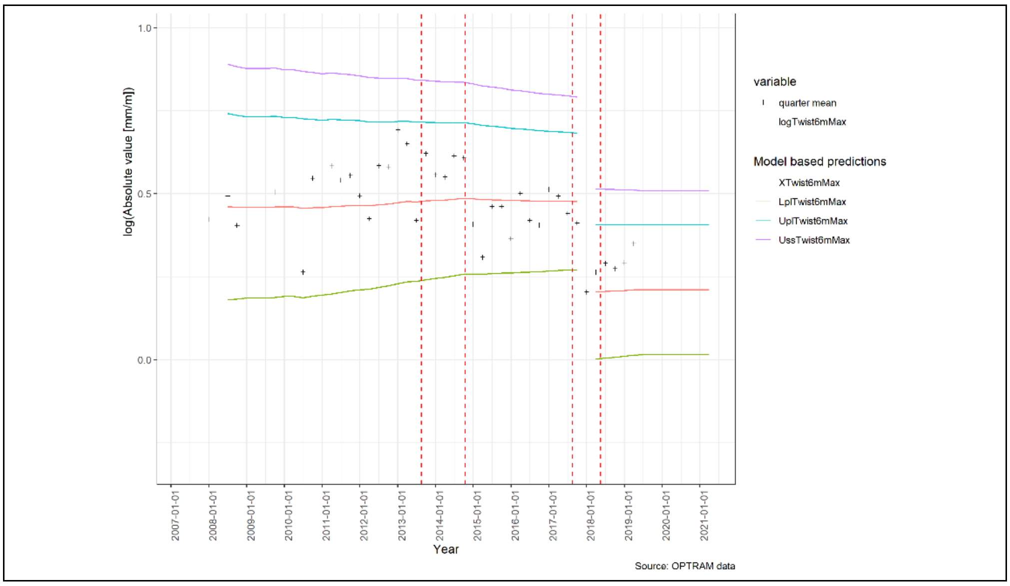

Figure 3. Time series model of the 3m twist variable for a 200 m segment obtained from track section 119 between 2007 and 2019.

Graphical condition assessment and prognoses

Graphics that are pictorial representations of data have the inherent property that they can make complex phenomena more easily interpretable (see, e.g., Bergquist & Söderholm, 2012, 2015). We will here discuss pictorial representations of data in the form of 2D heatmaps, and use colours to represent dimensions beyond the 2D imagery. Figure 4 shows a heatmap of data representing the condition of the track twist.

Figure 4 shows railway twist, here the 6m twist. The vertical axis represents the spatial information, and the horizontal axis represents the temporal information. The colours of the heatmap denote the twist values. The plot was created in R, using the ggplot2 package and the geom_tile command. The colours appear as streaks of mostly orange or yellow from left to right. The figure represents quarterly averaged data from ten years. The horizontal segmentation is a representation of 200 m segments of the track section. The data for the maximum twist of the segment constitute the foundation for the model output, and the maximum twist values are those that are essential for safety concerns. Figure 4 does not show maintenance actions, but all segments were tamped at least three times.

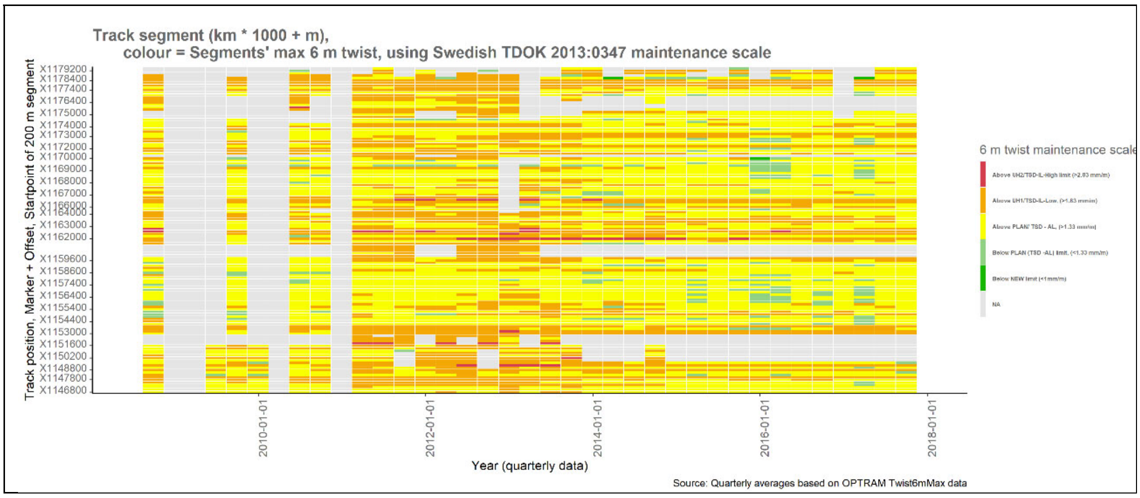

Figure 4. Graphical representation of track section 119 twist (6m) data from 2008-01-01 to 2018-03-31. The vertical axis represents the spatial dimension of the track, and the horizontal represents the temporal.

The observations have been segmented into bins representing maintenance classes, where dark green represents segments and times where the maximum 6 m twist for the segment was lower than the stipulated maximum for new tracks. Many conditions are worse than new, but they are still not of any concern (below the PLAN limit). Worse still are segments that are above the PLAN limit, which means that the maintenance organisation should consider maintaining the segment. The plot also shows even further degradations. When the track conditions have surpassed the UH1 (maintenance 1) limit, the maintenance organisation is required to plan maintenance so that they have performed it before the condition has reached the next level, UH2 (maintenance 2). If the condition exceeds the UH2, the maintenance organisation is required to maintain immediately. If the condition surpasses the KRIT (critical) limit, the track manager needs to close the track or impose speed restrictions until conditions have improved.

It is possible to add another layer to the plot, showing the performed maintenance, and Figure 5 shows the plot with a maintenance layer added. In this figure, it is clear that problematic segments usually remain as problematic even after tamping, since the colour remains the same.

The use of a geographical map for localisation purposes is attractive for linear assets such as railway track, but also for localising point assets (e.g. switches and crossings, bridges, and level crossings) as part of the infrastructure. Hence, a combination of heatmaps and geographical maps are useful for maintenance planning purposes. The use of GIS-applications and layers can also be used to integrate further information about the track surroundings that are useful for diagnostic and prognostic purposes, e.g. soil and water conditions. Furthermore, data about the weather should be useful for diagnostic and prognostic purposes. One example is the temperature, where low temperatures increase the risk for rail break and high temperatures increase the risk of buckling.

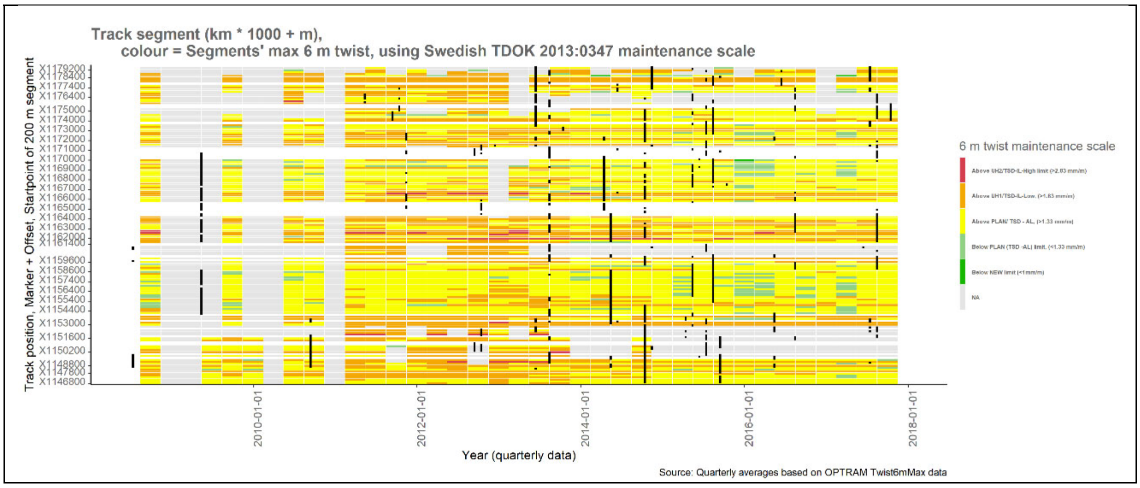

Figure 5. The image properties correspond to Figure 1, but here, the maintenance actions are also added (vertical black lines).

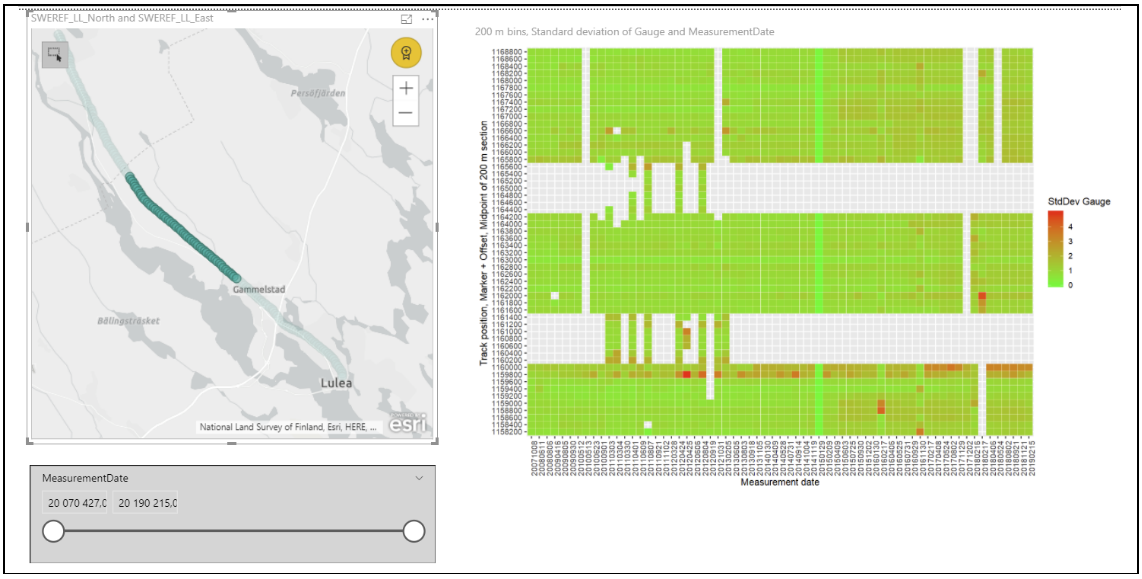

Figure 6. Power BI Desktop colourmap representation of, in this case, the gauge variation, where red segments vary the most.

Microsoft Power BI Desktop can also use R scripts and allows the user to run a collection of R packages for graphical presentations and analyses. This feature also allows for zooming, and Figure 6 shows an interactive display, allowing the user to zoom in on particular periods and track segments. In this plot, the colours are auto-scaled, from green in the colourmap representing segments with little gauge variation, to red where the variation is considerable. The Plotly package in R can also produce an interactive plot.

Another source of information that should be valuable when performing diagnostics and prognostics is the actual maintenance actions done on a segment (Arasteh Khouy, 2013; Famurewa, 2013). However, when dealing with the basic maintenance contracts, this information is stored locally within each contract (often in scanned PDF within a Share point solution) and not easily accessible for aggregated analysis. Appropriate usage of a standardised maintenance system (e.g. Maximo) should enable a more comprehensive analysis of performed maintenance actions (see, e.g., Al-Chalabi, 2018). The inspection system Bessy (e.g. inspection occasions and inspection remarks), the fault reporting system 0Felia (e.g. faults, repair times, and repair actions) also contain some useful additional information about maintenance actions and the asset condition. The asset register (BIS) also contains some information that might be useful for analysis purposes, e.g. type of items and installation year. The relevant age of infrastructure items can be calculated by a combination of the calendar time and the yearly tonnage.

Contributions

The datasets that the measurement trains produce are large and riddled with defects, and things to consider for sifting out the relevant results are presented. The graphical plots show that some maintenance actions are not effective, that there are underlying factors that make some track segments to outperform others, that continue to be troublesome despite repeated maintenance actions. Such sections that have been troublesome for years may eventually need non-standard corrective maintenance or reinvestment stand out in the spatiotemporal graphical overview. The findings show that a proper analysis may reveal how much a maintenance action improves the asset condition, not only directly after the action, but also if the long-term deterioration rate is affected. Hence, both the infrastructure manager and the contracted entrepreneur can share a common view of the asset's condition and maintenance effectiveness as a basis for continuous improvement.

Acknowledgements

The authors would like to thank the Swedish Innovation Agency (Vinnova) and the Swedish Transport Administration (Trafikverket) for their financial support.

References

Al-Chalabi, H. S. (2018). "Life cycle cost analysis of the ventilation system in Stockholm's road tunnels", *Journal of Quality in Maintenance Engineering*, Vol. 24, No. 3, pp. 358-375.

Arasteh Khouy, I. (2013). Cost-effective maintenance of railway track geometry: a shift from safety limits to maintenance limits (Doctoral dissertation, Luleå University of technology).

Arasteh Khouy, I., Larsson-Kråik, P. O., Nissen, A., Juntti, U., & Schunnesson, H. (2014). Optimisation of track geometry inspection interval. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 228(5), 546-556.

Bergquist, B., & Söderholm, P. (2012). "Control charts for assessment of linear asset condition using both temporal and spatial information", In *International Workshop and Congress on eMaintenance: 12/12/2012-14/12/2012* (pp. 145-152). Luleå University of technology, Luleå, Sweden.

Bergquist, B. & Söderholm, P. (2015). Control Charts supporting Condition-Based Maintenance of Linear Railway Infrastructure Assets. *Journal of COMADEM,* Vol. 18, No. 2, pp. 7-16.

Bergquist, B. & Söderholm, P. (2016). "Predictive Modelling for Estimation of Railway Track Degradation", in Kumar, U. Ahmadu, A., Verma, A.K., Varde, P. *Current Trends in Reliability, Availability, Maintainability and Safety*, Springer International Publishing AG, Cham. pp. 331-347.

Famurewa, S. M. (2013). *Increased railway infrastructure capacity through improved maintenance practices* (Doctoral dissertation, Luleå tekniska universitet).

Karim, R., Söderholm, P., & Candell, O. (2009). Development of ICT-based maintenance support services. *Journal of Quality in Maintenance Engineering,* Vol. 15, No. 2, pp. 127-150.

Levrat, E., Iung, B., & Crespo Marquez, A. (2008). E-maintenance: review and conceptual framework. *Production Planning & Control,* Vol. 19, No. 4, pp. 408-429.

Patra, A. P., Söderholm, P., & Kumar, U. (2009). Uncertainty estimation in railway track life-cycle cost: a case study from Swedish National Rail Administration. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 223(3), 285-293.

Stenström, C. (2015). *Effektsamband för underhåll av järnväg: Förstudie*. Kungliga tekniska högskolan, Inst. för transportvetenskap. (In Swedish).

Soleimanmeigouni, I., Ahmadi, A., Arasteh Khouy, I., & Letot, C. (2018). Evaluation of the effect of tamping on the track geometry condition: A case study. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 232(2), 408-420.

Lukoianova, T., & Rubin, V. L. (2014). Veracity roadmap: Is big data objective, truthful and credible? *Advances In Classification Research Online,* Vol. 24, No. 1.