Prediction of railway track geometry defects: a case study

Iman Soleimanmeigouni, Alireza Ahmadi, Arne Nissen & Xun Xiao

To cite this article: Iman Soleimanmeigouni, Alireza Ahmadi, Arne Nissen & Xun Xiao (2019): Prediction of railway track geometry defects: a case study, Structure and Infrastructure Engineering

To link to this article: https://doi.org/10.1080/15732479.2019.1679193

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 22 Oct 2019.

Submit your article to this journal

View related articles

View Crossmark data
Prediction of railway track geometry defects: a case study

Iman Soleimanmeigounia, Alireza Ahmadi², Arne Nissen³ and Xun Xiaoc

¹Division of Operation and Maintenance Engineering, Luleå University of Technology, Luleå, Sweden; ²Trafikverket, Luleå, Sweden; ³School of Fundamental Sciences, Massey University, Palmerston North, New Zealand

ABSTRACT

The aim of this study has been to develop a data-driven analytical methodology for prediction of isolated track geometry defects, based on the measurement data obtained from a field study. Within the study, a defect-based model has been proposed to identify the degradation pattern of isolated longitudinal level defects. The proposed model considered the occurrence of shock events in the degradation path. Furthermore, the effectiveness of tamping intervention in rectifying the longitudinal level defects was analysed. The results show that the linear model is an appropriate choice for modelling the degradation pattern of longitudinal level defects. In addition, a section-based model has been developed using binary logistic regression to predict the probability of occurrence of isolated defects associated with track sections. The model considered the standard deviation and kurtosis of longitudinal level as explanatory variables. It has been found that the kurtosis of the longitudinal level is a statistically significant predictor of the occurrence of isolated longitudinal level defects in a given track section. The validation results show that the proposed binary logistic regression model can be used to predict the occurrence of isolated defects in a track section.

1. Introduction

Nowadays, railways are experiencing higher demands for the transportation of passengers and goods. This will in turn impose higher demands on the railway capacity and service quality. As a result, infrastructure managers are being compelled to make strategies and plans to meet new requirements, which include a higher level of resilience against failure, more robust and available infrastructure and cost reduction. To achieve these goals, one of the key elements is the employment of an effective and efficient maintenance programme. A large part of the railway maintenance burden concerns track geometry maintenance. Maintenance actions are used to control the degradation of the track and restore the track geometry condition to an acceptable state. Activities that can be applied in track geometry maintenance include manual interventions, tamping and stone blowing. Three indicators are used to represent the track geometry quality, i.e. the standard deviation over a specific track length, the mean value and the extreme value of isolated defects (EN 13848-5, 2008).

Track quality indices (TQIs) are mostly formed based on the standard deviation of the vertical and lateral track irregularities (Soleimanmeigouni, Ahmadi, & Kumar 2018). Since a TQI aggregates the track geometry measurements to represent the overall condition of track sections, it may not provide complete information about severe isolated defects in the track sections. Isolated defects are short irregularities in the track geometry that can dramatically increase the dynamic forces between the wheel and rail, which in turn will accelerate the growth or occurrence of internal rail defects. These defects have been classified by the Swedish Transport Administration, Trafikverket, into three severity levels, i.e. UH1 defects, UH2 defects and critical defects (Trafikverket, 2015). UH1, UH2 and critical defects are those defects which have exceeded the lower bound of the intervention limit, the upper bound of the intervention limit and the safety limit, respectively (see Figure 1 for details). The occurrence of severe geometry defects can cause comfort problems for passengers, damage to track components and an increase in the risk of derailment. Moreover, these defects cause unplanned maintenance activities on the railway track, which in turn decrease the track availability and safety and increase the maintenance cost. In order to develop a more efficient and effective maintenance plan, isolated defects must be considered in track geometry degradation modelling.

There are a limited number of studies that have considered isolated defects in their degradation modelling and maintenance planning. Arasteh Khouy, Larsson-Kråk, Nissen, Juntti, and Schunnesson (2014) used survival analysis and considered isolated defects of the longitudinal level, twist 3 m, and twist 6 m for the purpose of predicting the need for corrective tamping actions. They found that, in their case study, the Weibull distribution was the best fitted...
distribution for modelling the probability of the occurrence of geometry defects. Later Alemazkoor, Ruppert, and Meidani (2018) used survival analysis, taking into account a number of covariates, to predict the probability that a ‘yellow tag’ defect would turn into a ‘red tag’ defect in a given time. ‘Red tag’ defects can be defined as isolated defects that exceed the maintenance limits of the Federal Railroad Administration (FRA). ‘Yellow tag’ defects are those isolated defects which are as yet below the FRA limits, but will finally turn into ‘red tag’ defects (Alemazkoor et al., 2018). They chose the Weibull distribution for modelling the time of the occurrence of ‘red tag’ defects and modelled the scale parameter as a function of independent variables, i.e. initial absolute values of the amplitude and length of defects, the track code, the class of track, the operating speed and the tonnage. Later they used their model to predict the probability that a track section would contain at least one ‘red tag’ defect in a given time.

Andrade and Teixeira (2013) predicted the probability that a given track section would need unplanned maintenance due to a severe geometry defect. They applied logistic regression to predict that probability and considered the standard deviation of the longitudinal level and alignment, and the existence of a switch or bridge in a section as explanatory variables. He, Li, Bhattacharjiya, Parikh, and Hampapur (2014) proposed a degradation model for capturing the changes of the amplitude of geometry defects and for predicting the probability that a ‘yellow tag’ defect would turn into a ‘red tag’ defect in a given time. They considered four explanatory variables in their model, namely the monthly traffic, the total monthly number of cars, the total monthly number of trains and the number of inspections since the latest ‘red tag’ defect. In addition, they applied the Cox proportional hazard model to predict the probability of derailment by considering the number and the amplitude of ‘yellow tag’ defects. Cárdenas-Gallo, Sarmiento, Morales, Bolivar, and Akhavan-Tabatabaei (2017) proposed an ensemble classifier for predicting when a ‘yellow tag’ defect would turn into a ‘red tag’ defect. They studied the problem with regard to three different aspects, i.e. deterioration, regression and clustering. Regarding deterioration, they applied a gamma process to model the evolution of the amplitude of ‘yellow tag’ defects over time. To identify the relationship between the explanatory variables and the future state of defects, they applied a logistic regression model. In addition, they used support vector machines (SVMs) to predict the probability of the occurrence of a ‘red tag’ defect. Sharma, Cui, He, Mohammadi, and Li (2018) proposed a Markov decision process for track maintenance decision making considering geometry defects. They used three algorithms, i.e. a random forest algorithm, a SVM algorithm and a logistic regression algorithm to predict the occurrence of severe isolated defects.

Although the literature addressed the track geometry isolated defects, still there is a need to a comprehensive study on the evolution of isolated defect to identify their degradation pattern. Moreover, although the effect of tamping on standard deviation of geometry parameters are well studied in the literature, its effect on isolated defects has not been extensively studied (Soleimanmeigouni, Ahmadi, Arasteh Khouy, & Letot, 2018). The other important point that must be considered for track geometry degradation modelling is considering the occurrence of shock events. Shock events in degradation path are assumed to be negative, causing the increase in degradation rate and even the failure of the system. Furthermore, according to the reviewed papers, different approaches are used to predict the probability of occurrence of isolated defects by considering aggregated TQIs as explanatory variables. However, to the best of our knowledge there is no study concerning the relationship between kurtosis of geometry parameters and the probability of occurrence of isolated defects. To this end, the present study was undertaken to explore the prediction of isolated track geometry defects and to address the aforementioned issues.
The aim of the present study has been to model the track geometry degradation and to predict the occurrence of UH2 defects. In order to model the track geometry degradation, the evolution of the amplitude of the longitudinal level defects within a maintenance cycle was modelled using simple linear regression. In addition to gradual degradation, there can be an abrupt change in the degradation path in which the amplitude of the defect will dramatically increase over time. This phenomenon is called shock event, and this kind of event was also considered in our degradation modelling. Furthermore, the effectiveness of tamping intervention in rectifying the longitudinal level defects was analysed. In order to predict the probability of the occurrence of UH2 defects, a section-based model was developed using binary logistic regression, where the kurtosis of the longitudinal level was also considered as an explanatory variable.

In order to validate the model for the purpose of classification of track sections based on the presence of UH2 defects, the sensitivity and specificity of the developed model were calculated. Data obtained from the Main Western Line in Sweden were used for the purpose of model development and for the case study. The rest of the paper is organised as follows. Section 2 provides a background to the topics of track geometry parameters and maintenance limits. Section 3 presents information on the track line used for the case study and explains the data recording and preparation for the case study. Section 4 presents the process of developing the degradation model. The section-based model, which considers the relationship between common indicators for planned maintenance and the occurrence of UH2 defects, is presented in Section 5. Finally, Section 6 provides the conclusions and proposes directions for future studies.

2. Track geometry parameters and maintenance limits

Track geometry parameters are widely used to represent the track condition and to plan maintenance activities, and these parameters can be divided into five classes: longitudinal level, alignment, gauge, cant and twist. Longitudinal level is the geometry of the track centreline projected onto the longitudinal vertical plane. Alignment is the geometry of the track centreline projected onto the longitudinal horizontal plane. Gauge is the distance between the gauge faces of two adjacent rails at a given location below the running surface. Cant (cross-level) is the difference in height between the adjacent running tables computed from the angle between the running surface and a horizontal reference plane. Twist is the algebraic difference between two cross-levels taken at a defined distance apart, usually expressed as the gradient between the two points of measurement (SS-EN 13848-1: 2004 + A1, 2008). According to SS-EN 13848-1: 2004 + A1 (2008), for each geometry parameter the indicators provided in Table 1 should be used to represent the track geometry quality.

According to EN 13848-5 (2008) there are three limits for maintenance actions: the immediate action limit (IAL), the intervention limit (IL) and the alert limit (AL). If the IAL or safety limit is exceeded, there is a potential risk of derailment, and consequently a speed reduction or line closure must be imposed before conducting a corrective maintenance action. If the IL or corrective maintenance limit is exceeded, a corrective maintenance action is required before the IAL is reached. If the AL or preventive maintenance limit is exceeded, the track geometry must be analysed for the planning of future maintenance actions. The European standard EN 13848-5 (2008) provides the IALs, ILs and ALs for isolated defects and the ALs for standard deviations. Generally, track quality indicators based on the standard deviation of the track geometry parameters are used to plan and perform preventive maintenance actions. On the other hand, the execution of corrective maintenance actions is based on the severity of isolated defects. Whenever the amplitude of an isolated defect exceeds the IL or IAL, corrective maintenance should be conducted on the track. Figure 2 shows the different maintenance zones based on the aforementioned limits.

The IALs are normative and take into account the track–vehicle interaction and the risk of unexpected events, whereas the ILs and the ALs are informative and are mainly linked with the maintenance policy. The ILs and ALs provided in EN 13848-5 (2008) reflect the common practice of most European infrastructure managers. In alignment with the European standard EN 13848-5 (2008), Trafikverket has defined four main limits, namely the planning limit, the UH1 limit, the UH2 limit and the critical limit, as can be seen in Figure 1. In Trafikverket (2015), the intervention limit is expressed as a range rather than a discrete value. Track irregularities that exceed the UH1 limit must be assessed for conducting maintenance before the UH2 limit is exceeded. For track irregularities exceeding the UH2 limit,

<table>
<thead>
<tr>
<th>Geometry parameter</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal level</td>
<td>Isolated defects that exceed a prescribed threshold. The indicator is mean to peak value. Standard deviation over a defined length, typically 200 m, in the wavelength range D1.</td>
</tr>
<tr>
<td>Alignment</td>
<td>Isolated defects that exceed a prescribed threshold. The indicator is mean to peak value. Standard deviation over a defined length, typically 200 m, in the wavelength range D1.</td>
</tr>
<tr>
<td>Cross-level</td>
<td>Absolute value. Isolated defects that exceed a prescribed threshold. The indicator is zero to peak value. Standard deviation over a defined length, typically 200 m.</td>
</tr>
<tr>
<td>Twist 3 and 6 m</td>
<td>The identification of individual defects which exceed a prescribed threshold. The measured track gauge. The difference between the measured track gauge and the nominal track gauge. The mean track gauge over a specified distance. The variation of the track gauge over a specified distance.</td>
</tr>
<tr>
<td>Gauge</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. The indicators used for representing the track geometry quality.
maintenance action must be taken without unnecessary delay. Therefore, track irregularities must be corrected before the UH2 limit is reached. Table 2 shows the relation between the limits defined in EN 13848-5 (2008) and those defined in Trafikverket (2015).

Based on the aforementioned limits, geometry defects can be classified according to their severity into three groups, i.e. UH1, UH2 and critical defects. UH1, UH2 and critical defects occur when track irregularities exceed the UH1, UH2 and critical limits, respectively. Examples of these three categories of defects are shown in Figure 1.

### 3. Data collection and preparation

Track geometry data for line section 414 between Järna and Katrineholm Central Station, collected from January 2015 to July 2018, were used to perform the case study presented in this paper. Line section 414 is part of the Main Western Line in Sweden (Västra Stambanan). The line section data were taken from Optram, which is the system used by Trafikverket for the study of measurements performed on the track and overhead lines. The maximum speed of trains on the Main Western Line is around 200 km/h. Line section 414 is 82 km long and consists of UIC 60 and SJ 50 rails, M1 ballast, Pandrol e-Clip fasteners and concrete sleepers. The annual passing tonnage of the line section is around 20 MGT.

### 3.1. Data alignment

Generally, track geometry measurement data obtained by inspection cars suffer from measurement errors and positional errors. The positional error can be as much as 100 m in some cases (Wang et al., 2018). Positional errors negatively affect the accuracy of the track irregularity predictions achieved by track geometry degradation models. A main source of positional errors is the presence of error in reference milestones installed along the track or the Global Positioning System (GPS), while another source of positional error is errors in the measured travelling distance. The accuracy of the measured travelling distance is affected by various factors, including the condition of the wheel–rail contact (Xu et al., 2013). In order to correct the positional measurement data, absolute position-based (APB) and relative position-based (RPB) methods can be applied. APB methods correct the positional measurement data based on the estimated absolute position information of the physical reference position. RPB methods correct the positional measurement data according to the estimated positional shift relative to historical inspection data. RPB

---

**Table 2. Maintenance limits defined in the European standard EN 13848-5 (2008) and Trafikverket (2015).**

<table>
<thead>
<tr>
<th>EN 13848-5 limits</th>
<th>Trafikverket’s limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alert limit</td>
<td>Planning limit</td>
</tr>
<tr>
<td>Intervention limit</td>
<td>UH1 limit (lower bound for corrective maintenance)</td>
</tr>
<tr>
<td></td>
<td>UH2 limit (upper bound for corrective maintenance)</td>
</tr>
<tr>
<td>Immediate action limit</td>
<td>Critical limit</td>
</tr>
</tbody>
</table>

**Table 3. Descriptions of the line section used in this study.**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>length</td>
<td>82 km</td>
</tr>
<tr>
<td>Number of track sections</td>
<td>440</td>
</tr>
<tr>
<td>Ballast type</td>
<td>M1</td>
</tr>
<tr>
<td>Rail type</td>
<td>UIC 60 and SJ 50</td>
</tr>
<tr>
<td>Sleeper type</td>
<td>Concrete</td>
</tr>
<tr>
<td>Fastener type</td>
<td>Pandrol e-Clip</td>
</tr>
<tr>
<td>Annual passing tonnage</td>
<td>20 MGT</td>
</tr>
<tr>
<td>Maximum allowable speed</td>
<td>200 km/h</td>
</tr>
</tbody>
</table>

---

**Image**

![Figure 2. Different maintenance zones based on limits.](image-url)
methods are used to align measurement data collected in different inspections (Xu, Sun, Liu, & Souleyrette, 2016). For more studies regarding the ABP methods and RBP methods, readers are referred to the work by Wang et al. (2018) and Xu et al. (2016).

Since the aim of our study has been to analyse the evolution of location-specific defects over time, the possession of accurate positions of the defects is of crucial importance. In order to align the measurement data of different inspections, OPTRAM uses a cross-correlation algorithm at specified intervals. Data alignment can be accomplished by cross-correlation analysis performed on the measured waveforms of inspections for different numbers of channels. The first channel roughly aligns the measurement data and the other channels refine the alignment (Selig, Cardillo, Stephens, & Smith, 2008). In the present study, the data alignment was performed separately using three different channels, i.e. short wavelength longitudinal level (wavelength 1–25 m), medium wavelength longitudinal level (wavelength 25–70 m) and gauge. If two of the alignments are within the same range, this is proof that the alignment is correct. The main advantage of using the medium wavelength longitudinal level is that it is not affected by tamping. However, at lower speeds the medium wavelength is not recorded and is not available for data alignment. By analysing the measurement data, it was observed that before data alignment the positional errors were as much as 30 m, whereas after data alignment the positional errors were in a range of 3 m. Figure 3 presents an example of the cross-correlation analysis applied on measurement data in Optram to correct positional errors.

Figure 3a depicts the deviations in the longitudinal level over a track section recorded on two different inspection dates before data alignment. As can be seen, the measurements recorded in the second inspection do not represent the same positions as those recorded in the first inspection. By performing a cross-correlation analysis between the two measured waveforms, the position of the maximum value of the cross-correlation indicates whether the sample leads or lags. Figure 3b shows the cross-correlation between the two measured waveforms, $C_{21}$. By considering the location of the maximum value of the cross-correlation, it can be inferred that the location of the waveform measured in the inspection performed in February 2017 leads the location of the waveform measured in the inspection performed in February 2016 by 66 sample intervals. By considering the fact that the sampling interval for the IMV200 measurement car is 25 cm, there is a shift between the two measured waveforms corresponding approximately to $66 \times 0.25 = 16.5$ m. Therefore, to deal with the position shifting error, the data must be aligned by shifting the measurements along the track position as shown in Figure 3c.

### 3.2. Identification of tamping cycles

In order to model the track geometry degradation, the time and position of the tamping interventions conducted on the track should be extracted to identify the maintenance cycles. Later this information will be used to model the evolution of the amplitude of the longitudinal level defects within a maintenance cycle. The information about the time and position of the tamping actions performed on the track is extracted from the BIS database. BIS is an asset register database which contains information on infrastructure and facilities, agreements and the history of tamping and grinding. However, in a number of cases the tamping...
interventions are not registered in the BIS system, and for these cases Trafikverket’s experts use certain criteria to identify the unregistered tamping actions.

Figure 4 displays the tamping zone based on the standard deviation of longitudinal level before tamping and the tamping ratio for the identification of unregistered tamping interventions. Track sections with a standard deviation of longitudinal level less than 80% of the UH1 limit are considered as good track sections, and track sections with a standard deviation of longitudinal level larger than 80% of the UH1 limit are considered as poor track sections. For a good track section, a 15% reduction in the standard deviation of longitudinal level is considered as a tamping intervention. For poor track sections, Equation (1) is used to identify unregistered tamping interventions:

\[ TR = \frac{SDLL_a}{SDLL_b} < 0.9 - \frac{0.16}{SDLL_b}, \]  

where \( TR \) is the tamping ratio, \( SDLL_a \) is the standard deviation of longitudinal level after tamping, and \( SDLL_b \) is the standard deviation of longitudinal level before tamping. A tamping intervention was conducted on the track section if the tamping ratio satisfies the inequality in formula (1).

By applying the mentioned criteria and using information from the BIS database, the positions and times of the tamping interventions conducted on line section 414 are identified. Figure 5 illustrates the variation of the standard deviation of longitudinal level over time for different track sections in line 414. By considering that a reduction in the evolution path of the standard deviation of longitudinal level represents a tamping intervention, Figure 5 can be used to identify the time and position of tamping interventions.

By considering the maintenance limits, Figure 5 can be used for an overall evaluation of the track geometry quality of the line section. Trafikverket (2015) stipulates that for the speed class of line section 414 the planning limit and the UH1 limit for the standard deviation of longitudinal level are 1.15 mm and 1.6 mm, respectively. Figure 5 shows that most of the track sections have a standard deviation of longitudinal level below the UH1 limit over the time period of the study. In addition, it can be seen that line tamping was not performed on line section 414 in the period of the study.

4. Modelling the evolution of geometry defects

The track geometry parameter which usually drives the need for maintenance activities is the short-wavelength longitudinal level irregularities (UIC, 2008). In order to model the track geometry degradation, the evolution of the amplitude of longitudinal level defects was studied.

4.1. Degradation model

The standard deviation of longitudinal level is widely used to represent the overall quality of railway track and to plan preventive maintenance actions. However, this aggregated indicator will not provide detailed information about the track condition and the severity of isolated defects in a track section. Therefore, there is a need to monitor and analyse the changes in the amplitude of isolated defects over time, to prevent the occurrence of UH2 defects in a track section. Figure 6 is a heat map of longitudinal level measurements at each sampling interval (every 25 cm) over time and within a given track section. Figure 6 can be used to observe the evolution of longitudinal level defects over time and to find the positions prone to the occurrence of UH2 defects. As can be seen, there are a few positions for which an increasing trend in the amplitude of the defects can be observed. These positions are prone to the occurrence of UH2 defects and must be considered for further analysis.

In order to detect the positions prone to the occurrence of UH2 defects, the six measurements with the highest values and the six measurements with the lowest values are recorded in Optram in each inspection run for each section and for both the right rail and the left rail. Those defects which have exceeded the planning limit may turn into UH2 defects in a short period of time and must be considered carefully for planning maintenance activities. Therefore, the changes in the amplitude of these defects over time must be analysed. The trend of the changes in the amplitude of the defects can be used to predict when a defect which has exceeded the planning limit will turn into a UH2 defect. The positions of defects are checked in each inspection run and defects within a distance of 3 m from each other are considered as belonging to the same defect. Figure 7 displays the evolution of the amplitude of the two longitudinal level defects at the positions 59.050 km and 59.074 km.

In order to identify the evolution pattern of the amplitude of the defects in line section 414, the same type of plot as that used for Figure 7 was used to provide plots for all the longitudinal level defects in this case study. By analysing the changes in the amplitude of the longitudinal level defects, it was observed that the degradation within a maintenance cycle had a linear pattern. Therefore, the linear regression model presented in Equation (2) was applied to model the evolution of isolated longitudinal level defects:

\[ A(t) = A_0 + \beta(t-t_{tamp}) + \epsilon, \]  

where \( A(t) \) is the absolute value of the amplitude of the longitudinal level defect in time \( t \), \( A_0 \) is the absolute amplitude of longitudinal level defect after the latest tamping intervention, \( \beta \) is the degradation rate and \( t_{tamp} \) is the latest tamping
time, as shown in Figure 8. Finally, $\varepsilon$ is the Gaussian random error term with a mean value equal to zero:

$$\varepsilon \sim N(0, \sigma^2)$$  \hspace{1cm} (3)

By using the measurement data recorded by the inspection cars, the parameters of the proposed model are estimated. When the amplitude of a defect exceeds the planning limit, the proposed model can be used to predict the time of the occurrence of a UH2 defect. In order to check the normality assumption for the residuals of the simple linear model, the Kolmogorov–Smirnov (KS) test is applied. The results of the KS test applied in the present study are summarised in Figure 9 in a histogram of the $p$ values. As can be seen, all the $p$ values of the KS test were larger than the significance level ($0.05$). Consequently, it could be concluded that the normality assumption for the residuals of the simple linear model was suitable for our case study.

Figure 10 illustrates the distribution of the degradation rates with a histogram of the degradation rates. It can be observed in this figure that most of the defects have a very small degradation rate, but the tail of the histogram indicates that there are a number of defects with a high degradation rate. These defects must be monitored and analysed as they may turn into UH2 defects in a short period of time.

In order to analyse the effect of the presence of a special asset in a track section, e.g. a switch and crossing, on the degradation rates, a box plot was created, as presented in Figure 11. Figure 11 shows that defects located in a track section with a special asset have a higher degradation rate on average. Therefore, when an isolated defect exceeds the planning limit in these sections, special consideration should be given to them.
4.2. Effect of shock events on the degradation path

By analysing the degradation path of geometry defects in line section 414, it is observed that in addition to gradual degradation, there can be an abrupt change in the degradation path in which the degradation level dramatically increases over time. This phenomenon is called shock event. Although the occurrence of an abrupt change in the degradation path is not common and may only occur in a small number of sections, considering them when modelling track geometry degradation is very important (Soleimanmeigouni et al., 2018). Track sections with an unusual trend in the degradation path normally have a significantly higher degradation rate after a shock event, causing a shorter maintenance cycle compared to that of track sections with a normal degradation path. Shock events in the degradation path may cause safety problems and, in the worst-case scenario, lead to derailment.

To analyse the occurrence of shock events in the degradation path, the changes in the evolution of the standard deviation of the longitudinal level and amplitude of the longitudinal level defects over time are monitored and analysed. Figure 12a depicts the changes in the standard deviation of longitudinal level for a track section with a shock event. When there is a shock event in the degradation path of the standard deviation of longitudinal level, there is also a

![Figure 7. Changes in the amplitude of isolated defects of longitudinal level.](image)

![Figure 8. Track geometry degradation parameters.](image)

![Figure 9. Histogram of the p values for the normality of the residuals using the KS test.](image)

![Figure 10. Histogram of the degradation rates.](image)

![Figure 11. Box plot of the degradation rates for defects located in sections with and sections without a special asset.](image)
change point in the degradation path of the isolated defect (see Figure 12b).

Figure 13 illustrates in detail the occurrence of a shock event by showing the longitudinal level measurements obtained in the 5th, 6th and 7th inspection of the track section concerned in Figure 12. As shown in Figure 13, there is a dramatic increase in the amplitude of the two isolated defects. The defect amplitude was below the planning limit in the 5th inspection, while it turned into a UH2 defect in the 7th inspection. Therefore, it can be concluded that the reason for the unusual trend observed in the evolution of the standard deviation of the longitudinal level was a problem that happened at the location within the track section marked with red cycle in Figure 13.

Shock events in the degradation path can occur for different reasons. For a number of track sections that have a change point in the degradation path, it has been observed that the increase in the defect amplitude occurred at exactly the same position as where a sleeper was replaced or a steel drum was installed. If after a sleeper replacement the ballast is not compacted well, this causes a change point in the track geometry degradation path. In addition, an inadequate installation or replacement of steel drums may also cause some damage to the track geometry condition and result in a change point in the degradation path. In the event of a shock event occurring in the degradation path, the regression model presented in Equation (2) should be slightly modified to model the track geometry degradation as follows:

\[ A(t) = A_0 + \beta(t-t_s) + \epsilon, \quad (4) \]

where \( t_s \) is the time of the occurrence of the shock event. Since after a shock event there is a new degradation pattern with a higher degradation rate, one must only use the data recorded after the shock event to model the evolution of the amplitude of the defects. After identifying a change point in the degradation path, the linear model expressed in Equation (4) can be used to predict the time of the occurrence of UH2 defects.

4.3. Effect of tamping on isolated defects

The effect of tamping on the degradation of defects whose amplitude has exceeded the planning limit can be seen by a sudden change in their amplitude after a tamping intervention, as is indicated by the vertical black dashed line in Figure 14. This figure illustrates in detail the effect of tamping on the amplitude of isolated defects by showing the longitudinal level waveform of a given track section before and after a tamping intervention. Figure 14a–c represents the longitudinal level waveform before the tamping intervention, one month after the tamping intervention and one year after the tamping intervention, respectively. As can be seen in Figure 14a, there is a UH2 defect in the track...
Figure 13. Unusual growth of the amplitude of isolated longitudinal level defects.

Figure 14. Evolution of longitudinal level irregularities for a single track section.
section before the tamping intervention. Figure 14b shows that the tamping intervention rectified the UH2 defect and all the measurements are within the planning limits one month after the tamping intervention. As can be seen in Figure 14c, although the tamping rectified the UH2 defect, after one year a UH1 defect occurred at the same position.

This issue was analysed for all the track sections in line section 414 and it was observed that for around 35% of the sections with a UH1 defect or UH2 defect, after a tamping intervention a UH1 defect or UH2 defect occurred again at the same position. This shows that any rectification of isolated defects through the current maintenance practice is not durable. In fact, correcting isolated defects using tractors and lightweight machines or tamping machines cannot remove the root causes of defects. Many problems in the track can be considered as root causes of isolated defects, e.g. broken sleepers, track substructure problems and drainage problems.

5. Section-based model

The proposed degradation model can be used to predict the time of the occurrence of UH2 defects. However, generally railway infrastructure managers prefer to plan maintenance activities based on overall condition of track sections rather than isolated defects. Moreover, generally TQIs based on the standard deviation of longitudinal level is used for planning track geometry maintenance activities. Therefore, a section-based model is proposed which considers the standard deviation and kurtosis of longitudinal level to predict the probability of the occurrence of UH2 defects in a given period of time.

5.1. Binary logistic regression

Regression is a well-known statistical method in data analysis for describing the relationship between a response variable and a set of explanatory variables (Hosmer & Lemeshow 2000). However, in many cases the response variable is binary or dichotomous. Modelling a binary response variable using a linear regression which assumes that the response is normally distributed will result in biased parameter estimates. In the case that the response is a binary variable the logistic regression which links the probability of a binomial distribution to the explanatory variables after a suitable transformation has been used as a standard method.

By considering \( Y \) as the binary response variable, the aim is to model the conditional probability \( P(Y = 1|x) = \pi(x) \) as a function of a set of explanatory variables. In logistic regression, the logit transformation of the probability \( \pi(x) \) is modelled as the linear function of the explanatory variables, as follows:

\[
g(x) = \logit(\pi(x)) = \log \left( \frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n,\]

where \( n \) is the number of explanatory variables, \( x = (x_1, \ldots, x_n) \) are explanatory variables, \( \beta_0, \beta_1, \ldots, \beta_n \) are the model coefficients, the logit denotes the logit transformation, \( \logit(\pi(x)) \), is the odds of \( Y \) and \( g(x) \) is the linear function associated to the logit model. By solving the logit equation with respect to \( \pi(x) \), the probability of the occurrence of an event \( P(Y = 1|x) \) can be obtained using:

\[
P(Y = 1|x) = \pi(x) = \frac{\exp(g(x))}{1 + \exp(g(x))}.
\]

In order to estimate the coefficients of the model, the maximum likelihood method can be applied. Given \( m \) observations as \((x_j, y_j), j = 1, \ldots, m\), the likelihood function is given by:

\[
L(\beta_0, \beta_1, \ldots, \beta_n) = \prod_{j=1}^{m} \pi(x_j)^{y_j}(1 - \pi(x_j))^{1-y_j}.
\]

The estimation of the model coefficients \( \beta_0, \beta_1, \ldots, \beta_n \) will be obtained by maximisation of the likelihood function or equivalently the log-likelihood function.

5.2. Relationship between the standard deviation and kurtosis of longitudinal level and the occurrence of UH2 defects

The standard deviation is used to show the variation or dispersion of track geometry measurement data. A low standard deviation indicates that the geometry measurements are close to the mean and a high standard deviation indicates that the geometry measurements have a high variation around the mean value. Since the presence of extreme values in the geometry measurements is of importance, then kurtosis can provide useful information. Kurtosis is a measure of tailedness of data. Higher kurtosis is the result of infrequent extreme observations (or outliers), as opposed to frequent moderate deviations. Therefore, it is expected that a track section where most of the geometry measurements are low, but which contains a UH2 defect, will have a high kurtosis.

Figure 15 presents the relationship between the standard deviation and the kurtosis of the longitudinal level with the presence of a UH2 defect in the track section. The left panel of Figure 15 shows the waveforms of the longitudinal level for three track sections in the same kilometre of line section 414. The right panel of the figure shows the estimated density of the longitudinal level measurements of the track sections obtained with a kernel smoothing function. In both track sections ‘a’ and ‘b’ there is a UH2 defect. However, in section ‘a’ the longitudinal levels of most of the sample points are below the planning limit, whereas in section ‘b’ there are a number of defects which have exceeded the planning limit. As a result, section ‘a’ has a smaller standard deviation than section ‘b’, but has a higher kurtosis than section ‘b’. This point is clear from the density function of the two waveforms in that the density function of section ‘a’ has a sharper peak and longer tails than that of section ‘b’.

When studying the longitudinal level waveform of section ‘c’ in Figure 15, one can observe that all the longitudinal level measurements are below the planning limit and there
is no UH2 defect in that section. As expected, the density function for this section has a sharper peak and shorter tails than that for the other sections, which indicates that section ‘c’ has a smaller kurtosis than the other two. In addition, the density functions of sections ‘a’ and ‘b’ are flatter than the density function of section ‘c’, which indicates that section ‘c’ has a smaller standard deviation.

In order to find out how standard deviation and kurtosis of longitudinal level are related to the occurrence of UH2 defects, the binary logistic regression was applied. The response variable takes a value of 1 with the occurrence of at least one UH2 defect, and otherwise it takes a value of 0. The results for the fitting of the model are summarised in Table 4.

In this study, the statistical significance of the individual regression coefficients was tested using the chi-square statistic. As can be seen in Table 4, the \( p \) values of all the predictors are less than the significance level. Therefore, it can be concluded that all the variables in the model have a significant effect on the probability of the occurrence of UH2 defects. The Hosmer–Lemeshow (H-L) test was applied to assess the fit of the applied logistic regression model against actual outputs. A \( p \) value for the H-L test less than the significance level (\( p \) value < 0.05) indicates a poor fit to the data and leads to the conclusion that the predicted probabilities from the logistic regression model deviate from the observed proportion of the events. The value obtained for the H-L goodness-of-fit statistic was 10.63 and the corresponding \( p \) value from the chi-square distribution with eight degrees of freedom was 0.23.

Therefore, it can be inferred that the proposed model was properly fitted. According to Table 4, both the standard deviation and the kurtosis of longitudinal level are statistically significant and have positive coefficients. This means that the higher the standard deviation and kurtosis of longitudinal level is, the higher is the probability of the occurrence of UH2 defects. This finding is illustrated in Figure 16, which shows the relationship between the standard deviation and kurtosis of longitudinal level measurements and the presence of UH2 defects in the whole of line section 414.

Figure 16 shows that when the standard deviation or the kurtosis of longitudinal level is low, there is no UH2 defect in the track section in question. Moreover, whenever the standard deviation of the longitudinal level is higher than the UH1 limit and the kurtosis is low (close to zero), a very small number of UH2 defects have occurred. Similarly, when the kurtosis is high and the standard deviation is low, a small number of UH2 defects have occurred. Therefore, both the standard deviation and the kurtosis of the longitudinal level must be higher than some specified value for a UH2 defect to be occurring in the track section. Therefore, considering both the standard deviation and the kurtosis as the explanatory variables in predicting the occurrence of UH2 defects may increase the prediction accuracy.
5.3. Prediction of the occurrence of UH2 defects

In order to predict the probability of the occurrence of UH2 defects in a track section in a given period of time, binary logistic regression was applied to develop the section-based model. The response variable $Y$ takes a value of 1 whenever there is at least one UH2 defect in the track section and a value of 0 whenever the section only contains defects which have exceeded the planning limit or the UH1 limit. In the section-based model, the aim is to predict the probability that a section which contains defects which have exceeded the planning limit or the UH1 limit will turn into a section containing at least one UH2 defect in a given period of time. The explanatory variables considered in the model are the following four variables: (1) the standard deviation of longitudinal level, (2) the kurtosis of longitudinal level, (3) the presence of defects which exceeded the planning limit or the UH1 limit in the latest measurement and (4) the time interval (in years). The binary logistic regression for the section-based model is as follows:

$$P(Y = 1) = \frac{e^{\beta_0 + \beta_1 SDLL + \beta_2 \gamma_2 + \beta_3 \Delta + \beta_4 c}}{1 + e^{\beta_0 + \beta_1 SDLL + \beta_2 \gamma_2 + \beta_3 \Delta + \beta_4 c}},$$

where $SDLL$ is the standard deviation of longitudinal level, $\gamma_2$ is the kurtosis excess of longitudinal level, $\Delta$ is the time interval and $c$ is a categorical variable which takes a value of 0 if the section only contains a defect which exceeded the planning limit in the latest measurement and a value of 1 if the section also contains at least one UH1 defect. Maximum likelihood estimation method is used to estimate the model parameters. The key results for the fitting of binary logistic regression to the data are summarised in Table 5.

As can be seen in Table 5, the $p$ values of all the predictors are less than the significance level. Therefore, it can be concluded that all the variables in the model have a significant effect on the probability of the occurrence of UH2 defects. The value obtained for the H-L goodness-of-fit statistic was 7.21 and the corresponding $p$ value from the chi-square distribution with eight degrees of freedom was 0.514. Therefore, it can be inferred that the proposed model is properly fitted. According to Table 5, the coefficient of time is positive, which means that the probability of the occurrence of a UH2 defect in a track section is higher in a longer period of time. In addition, the coefficient of the categorical variable $c$ is positive. This means that when a track section has been found to have a UH1 defect in the latest measurement, the probability of the occurrence of a UH2 defect in the time period $\Delta t$ is higher than when the section only contains a defect which has exceeded the planning limit. In addition, the standard deviation and kurtosis of longitudinal level have a positive coefficient. This means that the higher the standard deviation and kurtosis are, the higher is the probability of the occurrence of a UH2 defect in the time period $\Delta t$.

The probabilities predicted using binary logistic regression can be used to classify the outputs by using a single cut-point ($\omega$) to compare each estimated probability with respect to $\omega$. If the estimated probabilities are greater than $\omega$, the predicted class will be equal to 1, $\hat{Y} = 1$, and otherwise the predicted class will be equal to 0, $\hat{Y} = 0$. In order to establish the final prediction, an optimal cut point value is estimated using the sensitivity and specificity. When classification is the main goal of the analysis, the sensitivity and specificity can be used to assess the model performance ( Hosmer & Lemeshow, 2000). The sensitivity and specificity are statistical measures of the performance of a binary classification test. The sensitivity or true positive rate (TPR) measures the proportion of correctly classified non-events ( Peng, Lee, & Ingersoll, 2002):

$$\text{Sensitivity} = \frac{\text{Number of true positives}}{\text{Number of true positives} + \text{Number of false negatives}}.$$
Specificity

\[
\text{Specificity} = \frac{\text{Number of true negatives}}{\text{Number of false positives} + \text{Number of true negatives}}
\]

In order to test the performance of the model using the available data, 70% of the data were used as a training set and 30% of the data were used as a test set. Sensitivity and specificity are dependent on the value of cut-point. Figure 17 presents the changes in sensitivity and specificity with respect to different cut-point values. Particularly, the sensitivity is monotonic decreasing in \( x \) while the specificity is monotonic increasing in \( x \) as shown in Figure 17. Therefore, it is impossible to maximise both of sensitivity and specificity simultaneously. In order to select the optimal cut-point, one should optimise a suitable measure which balances the sensitivity and specificity well. For example, the arithmetic mean of sensitivity and specificity is the simplest measure. One can also use a weighted average by incorporating different costs related to false positives and false negatives. Another widely used measure is the \( F \)-score which is the harmonic mean of sensitivity and specificity. In this regard, the sensitivity and specificity for different cut-points value are calculated and the cut-point which balances both of them, i.e. the sensitivity and specificity are equal, is selected to classify data.

As can be seen in Figure 17, the optimal value for the cut-point should be \( \alpha^* = 0.23 \), as this is the value at which the sensitivity and specificity curves cross each other. By considering \( \alpha^* \) as the optimal cut-point, the model sensitivity and specificity are 89%, which seems reasonable for this case study.

### 6. Conclusions

The aim of this study has been to develop a data-driven analytical methodology for the prediction of track geometry defects by performing an extensive case study on line section 414 of the Main Western Line in Sweden. In this study, particular emphasis has been placed on the prediction of UH2 defects, which entail great costs for the maintenance of railway tracks. In order to identify the degradation pattern of isolated defects, a detailed analysis on foot-by-foot track geometry measurement data has been performed. It is found that isolated longitudinal level defects have a linear degradation pattern. The modelling methodology considered the occurrence of shock events and their associated change points in the degradation path. Shock events in the degradation path may cause safety problems and, in the worst-case scenario, lead to derailment.

In addition, the effectiveness of tamping intervention in rectifying the longitudinal level defects was analysed. It was observed that for around 35% of the sections with a UH1 defect or UH2 defect, after a tamping intervention a UH1 defect or UH2 defect occurred again at the same position. This shows that in many cases, spot tamping using tractors and lightweight machines or tamping machines cannot remove the root causes of defects. In order to predict the probability of the occurrence of UH2 failures in a track section, binary logistic regression is applied in the model developed in this study. Four variables, namely the standard deviation and kurtosis of the longitudinal level, the time interval and the presence of defects in the latest measurement, are selected as the explanatory variables in the model. The results show that the probability of the occurrence of UH2 defects in a track section is linked to the standard deviation and kurtosis of the track section.

Therefore, it can be concluded that the aggregated TQIs are statistically significant predictors for the occurrence of UH2 defects. Using the proposed section-based model railway companies can predict which track section will need a maintenance due to the occurrence of a severe isolated defect. This will enable the railway companies to perform section-wise preventive maintenance. In addition, it is found that the kurtosis, which is a measure of the tailedness of the data, can be used efficiently to capture the information.
about the occurrence of UH2 defects. This is important because it affects the design and interpretation of future studies on TQIs. The applied binary logistic regression model shows a satisfactory performance in predicting the occurrence of UH2 defects, which further justifies our approach.

One possible future research direction would be to explore the prediction of other defect types, e.g. twist and alignment defects. Another possible direction would be to include more explanatory variables, e.g. the soil type, traffic information and the track type, in the logistic regression model to improve the prediction accuracy. Finally, a further research direction would be to combine the predictions produced through the degradation model and the section-based model to establish an integrated framework for efficient maintenance planning.

Notes

1. Underhåll 1 (in English: Maintenance 1)
2. Underhåll 2 (in English: Maintenance 2)
3. Baninformationssystem (in English: Track Information System)

Acknowledgements

The authors would like to thank Trafikverket, SIMTRACK project, and Luleå Railway Research Center (JVTC) for their technical and financial support during this project.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Xun Xiao http://orcid.org/0000-0002-8780-5471

References


