Construction and Building Materials 349 (2022) 128786

Contents lists available at ScienceDirect

Construction and Building Materials

journal homepage: www.elsevier.com/locate/conbuildmat

Big-data driven assessment of railway track and maintenance efficiency using Artificial Neural Networks

K. Popov [a,*], R. De Bold [a], H.-K. Chai [a], M.C. Forde [a], C.L. Ho [b], J.P. Hyslip [c], H.F. Kashani [d], P. Long [e], S.S. Hsu [f]

[a] University of Edinburgh, School of Engineering, Edinburgh EH9 3JL, UK
[b] University of Massachusetts Amherst, Amherst, MA 01003, USA
[c] HNTB Corporation, Parsippany, NJ 07054, USA
[d] Loram Technologies Inc, Georgetown TX, USA
[e] Loram UK Ltd, RTC Business Park, Derby DE24 8UP, UK
[f] Network Rail, One Stratford Place, London E20 1EJ, UK

ARTICLE INFO

Keywords:
Rail
Maintenance
High Speed Track
Ballast
Monitoring
Artificial Neural Network
Machine Learning
Tamping

ABSTRACT

This study will address the opportunity of using a big data driven approach to providing a more specific description of track quality, mainly selecting segments of the track exhibiting higher settlement with the use of data analytics and machine learning. The focus will be on a high-speed line in the UK with data covering over 15 years of track geometry. Data sets describing track geometry quality have an enormous volume, which means that it is impractical to apply conventional methods to process it fully. The overall aim of this work was to apply an AI technique to analyse the big data. An Artificial Neural Network (ANN) was developed on features from the available data set and used to identify segments of the track where the condition has either improved or deteriorated in the period between two inspection runs. The model achieved an accuracy of ∼ 98 % during training and was able to reliably identify segments of the track which underwent significant changes in measurements. The ANN was used to perform an initial analysis of the geometry data, which revealed that maintenance works (mainly large-scale tamping) may include healthy portions of the track and potentially reduce its life span. Approximately 50 % of the tamped track was not found to improve significantly, based on several different geometric features. On the other hand, local works with a sprinter tamper were much more efficient at eliminating defects.

1. Introduction

Railways are an important means of transportation in the UK, which has seen a steady increase in demand over time – Fig. 1. At the end of 2019–20 Q3, the recorded rail passenger journeys in the UK were 462 million, which was an increase of almost 12 million (2.6 %) from a year prior. Passenger revenue reached US$3.5 billion in Q3 – 2.9 % increase [1]. In June 2019, the UK committed to reaching a net zero carbon emission target by 2050 [2], which will ultimately increase rail traffic further. Currently, transportation is the largest contributor to greenhouse gas emissions in the UK [3] and if the goal set out is to be met, sustainable transportation services, such as trains, must become more reliable. The increasing use of railway transportation will lead directly to higher annual tonnage on tracks and greater care must be taken of the assets to ensure the safety and comfort of all passengers, as well as the cost effective and optimum maintenance of the track.

Rail tracks will deteriorate as they age, and this is a process which is usually affected by various factors. The deterioration may be in the form of deviations from track geometry, ballast fouling, loosening of fasteners, ties/sleepers cracking, wet beds, etc. and these faults can be the cause discomfort for passengers on-board a train, and in more extreme cases – derailments.

On high-speed routes there are strict regulations to maintain the high quality of the lines and avoid traffic disruptions, which can cause substantial financial losses. To do this, more rigorous and efficient ways of evaluating the condition of the tracks are necessary. It is reported that European countries allocate between US$20B and US$30B annually on maintenance and renewal of railway systems [4]. Maintenance is usually

* Corresponding author.
E-mail address: k.popov@ed.ac.uk (K. Popov).

https://doi.org/10.1016/j.conbuildmat.2022.128786
Received 8 June 2022; Received in revised form 9 August 2022; Accepted 11 August 2022
Available online 18 August 2022

in the form of track tamping. There are two kinds of tamping performed by Network Rail – large-scale and local. Large-scale (or high output) tamping deals with more severe faults, whereas local works (sprinter tamping or manual packing of the ballast) are against individual defects covering a short span of the track. The types of maintenance techniques will be further elaborated on later in this paper. Methods of estimating the severity and cause of track deterioration at a given point in time are crucial to the scheduling of proactive/predictive and not just corrective or preventive maintenance, which will extend the lifespan of railways. Since these are highly expensive activities, such methods could potentially save US$Ms annually.

This project will focus on a high-speed operated line in the UK. The aim of the paper is to provide a new method of evaluating track quality, and evaluate the efficiency of different maintenance activities. The objectives are as follows:

- Preliminary analysis of track geometry data to obtain domain knowledge
- Developing an Artificial Neural Network (ANN) to classify track segments based on on-going quality
- Assessment of maintenance efficiency based on various remediation activities.
- A rich dataset of geometry measurements from 2003 to 2021 has been obtained via Network Rail for this purpose.

2. Background

Some of the earlier methods used to estimate the degradation of railways over time were designed in the 1980s and '90s. These were mostly mechanistic models looking into defects caused by ballast settlement [5]. The main issue with mechanistic models has been their inability to consider factors of uncertainty of track degradation behaviour. The purpose of these was to understand how a track behaves, so predictions about its quality can be made. In more recent research, a numerical model was developed, which has the ability to calculate differential track settlement [6]. It is based on Finite Element Method with perfectly matched layers (FEM-PML) and solved across two domains (frequency-wavenumber and time–space). The model is used to update the track geometry profile after each axle load and assess its effect on differential track settlement. In [7], the authors developed a semi-analytical expression for track settlement at each load cycle. Provided a vehicle-track interaction analysis, the model was able to represent the differences in the rate of settlement due to differences in initial track bed stiffness, type and speed of vehicles, and the resulting rail roughness.

Apart from numerical models, a number of quality indices have been developed as a way of assessing the overall condition of a railway. Some of them are listed in [8] and include the Track Roughness index (developed by Amtrak), Track Geometry index (developed by Indian railways), J coefficient (Poland), and Track Quality index (Federal Railroad Administration - FRA). These are indirectly related to geometry measurements, as they are derived from their statistical distribution properties. They are thought to be more indicative of the roughness of the track, or its overall condition, rather than highlighting unique problems. While properties such as standard deviation (SD) provide information about the spread of a dataset, they do not inform the engineer about measurements at certain locations rapidly approaching critical levels. Without this information it is possible that there is inadequate allocation of tamping, leading to an inefficient scheme. Such practice can impact the long-term condition of the track, train safety, ride quality, and maintenance cost [9].

It is common practice, especially in Europe, to use the SD of a single geometric feature (rail profile, alignment, twist, etc.) as an index and set a threshold value to describe the quality of the track. This technique was used in [9], where different thresholds were simulated to optimize maintenance costs. The authors model track degradation and the probability of isolated defects based on SD levels from a segment of a fixed length. Other statistical models have been applied to assess quality as well and predict future behaviour using this kind of index [10–12]. Often, the length of track used to estimate SD is 200 m – this technique is one of two being used by Network Rail. Based on this, decisions where to tamp the track are made to ensure readings are within target values [13]. This kind of maintenance usually covers longer portions of the track and is known as preventive maintenance. The practice followed by Network Rail is similar to the one used by the Société Nationale des Chemins de fer Français (SNCF), the French rail operator. The other method of scheduling maintenance is based on individual readings exceeding predefined threshold values (corrective maintenance). This is usually aimed at removing local defects which have been flagged as warning values.

Structural health monitoring in civil engineering can vary depending on the type of asset being monitored. Data may be collected from static or moving sensors. In bridge monitoring, sensors will be fixed on different components of the structure and collect data continuously at a specific frequency [14,15] – these are static sensors. This creates a large data set in a time series format, where deep learning algorithms such as Long Short-Term Memory (LSTM) neural networks can be applied to predict future behaviour [16]. In the case of railway track infrastructure, data can be collected from fixed sensors, but this can be difficult due to

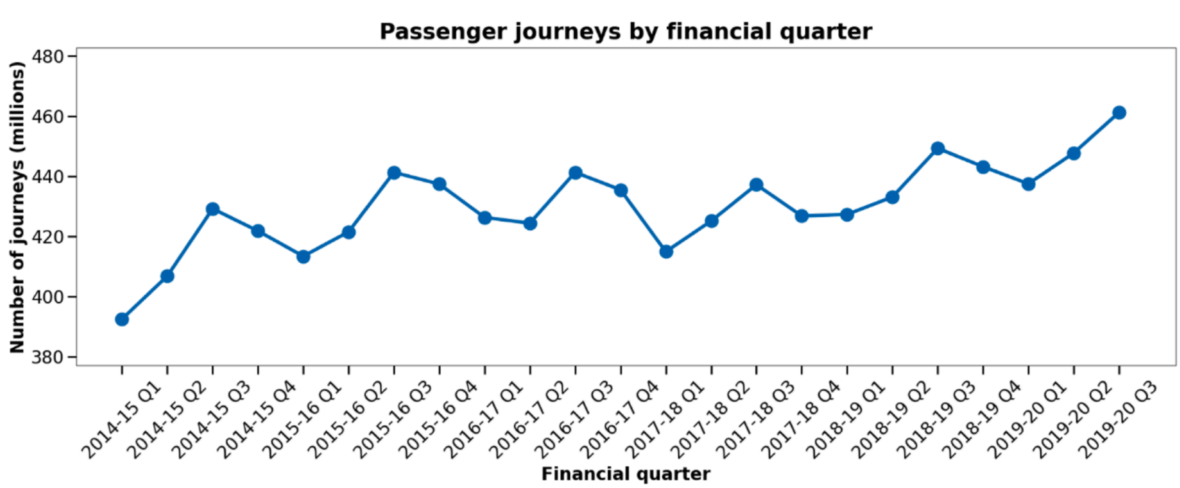

Fig. 1. Passenger journeys in the UK: 2014–2019. Data taken from [1].

the vast number of sensors required to cover an entire track. The quantity of geometry measurements generated from such a system would be computationally expensive and unfeasible. Instead, sensing equipment is fixed on a recording vehicle, which is used to collect data periodically. This is typically undertaken every few weeks. As a result, the data is less regular and requires a different technique to analyse. Various statistical methods have been applied to describe the state of a railway track based on geometry measurements [12,17–20]. Other probabilistic models, such as random forest regression have been used for predicting track degradation indices [21]. More traditional machine learning algorithms have also gained popularity in the railway industry as a way of analysing very large data sets in a relatively short period. [22,23]. Artificial neural networks have been used in a number of studies before and will be the main analysis tool in this paper as well.

An example of such an analysis is the research undertaken using track geometry measurements from the SNCF-operated high-speed TGV line in France [24,25]. In [25], the authors built a model based on a Monte Carlo simulation to represent the ageing process of the track under certain conditions. The main geometric feature used in this study is the mean deviation of the longitudinal rail profile (NL) of a 1-kilometre-long track section. The progression of this parameter in between tamping activities is simulated based on the assumption that the growth of degradation is described by an exponential function (Fig. 2):

$$NLinit_n = e^{b_n(t-t_n)} + \varepsilon(t) \tag{1}$$

where t is time, t_n is the time of the last tamping activity, b_n (deterioration rate) is a log-normally distributed stochastic variable, and $\varepsilon(t)$ is a normally distributed variable with mean equal to zero. $NLinit_n$ is the initial NL value immediately after the nth tamping and is log-normally distributed, i.e.

$$NLinit_n \; \mathscr{L}\mathscr{N}\left(\mu_{NLinit}(n), \sigma^2_{NLinit}(n)\right) \tag{2}$$

where $\mu_{NLinit}(n)$ is the mean and σ^2_{NLinit} – the variance. The factors affecting the track ageing process have remained the same throughout the simulation. The results showed that the $NLinit_n$ value grew with each tamping activity until it eventually stabilized to one value. This is indicative of the lower efficiency of tamping over time, i.e., the degrading state of the ballast. This was further noticeable since the quality loss rate (b_n) continued to diverge with each tamping, meaning that the condition is degrading quicker (Fig. 3). The developed model was used to simulate the longitudinal level of the high-speed TGV line given two different maximum allowable velocities (250 km/h and 300 km/h). The simulation showed that the two cases followed a very similar pattern, with the only difference being that the simulation with the higher velocity exhibited quicker deterioration and will require more frequent interventions.

A different study [26] has been carried out on UK railway lines regarding the whole lifecycle cost (WLCC) analysis. The work here mainly focuses on the economical aspect and analyses trade-offs between costs for transportation, construction, maintenance, and use. The track quality in this study was expressed as the standard deviation value of the horizontal and vertical geometry features over 200-metre segments. Monte-Carlo simulations were performed to generate a range of scenarios where the WLCC of an asset is estimated. One of the findings was that maintaining the quality of a high-speed passenger line at an average level was more beneficial economically. Intuitively, performing the minimum amount of tamping reduces the maintenance costs significantly, however that created a larger increase in track use costs, as well as increasing risk of derailment. The outcome of this study suggests that the maintenance scheduling of a railway track involves a trade-off between quality and cost and may be different for individual line classes. While this is an important aspect of quality assessment and maintenance scheduling, this study has not yet considered a lifecycle cost analysis and will focus mainly on assessing track quality.

3. Available data

3.1. Geometry measurements

The track investigated in this paper is strictly for high-speed trains and does not carry other types of traffic on its main lines. The majority of the line is ballasted track with twin and mono-block (mostly at switches and crossings) concrete sleepers. It is regularly inspected mainly using specialised track recording vehicles known as TRV's, but also the New Measurement Train (NMT) – Fig. 4. The NMT can collect readings from the track at up to 200 km/h. All recording vehicles are equipped with a variety of sensors (accelerometers, transducers, laser sensors, etc.) to measure specific geometric features and detect faults and degrading track. The trains will measure geometric properties, such as top profiles (over both 35 m (short) and 70 m (long) wavelengths), twist (3 m), gauge, alignment, and cant deficiency. This study will focus mostly on vertical profile related features. TRV's are used to inspect enormous lengths of track annually – 115,000 miles a year in the case of the NMT, producing large amounts of data making it difficult to analyse using traditional methods.

Readings from inspections on the line in question are stored on a server and access to these data have been granted to the University of Edinburgh for research purposes. The data set explored is from the period September 2003 – April 2021. Usually, inspections are carried

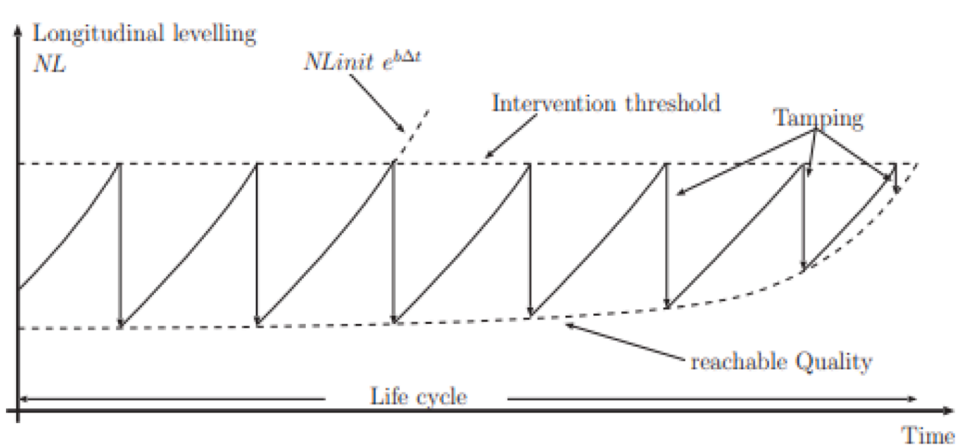

Fig. 2. Schematic track geometry quality course using an exponential model [25].

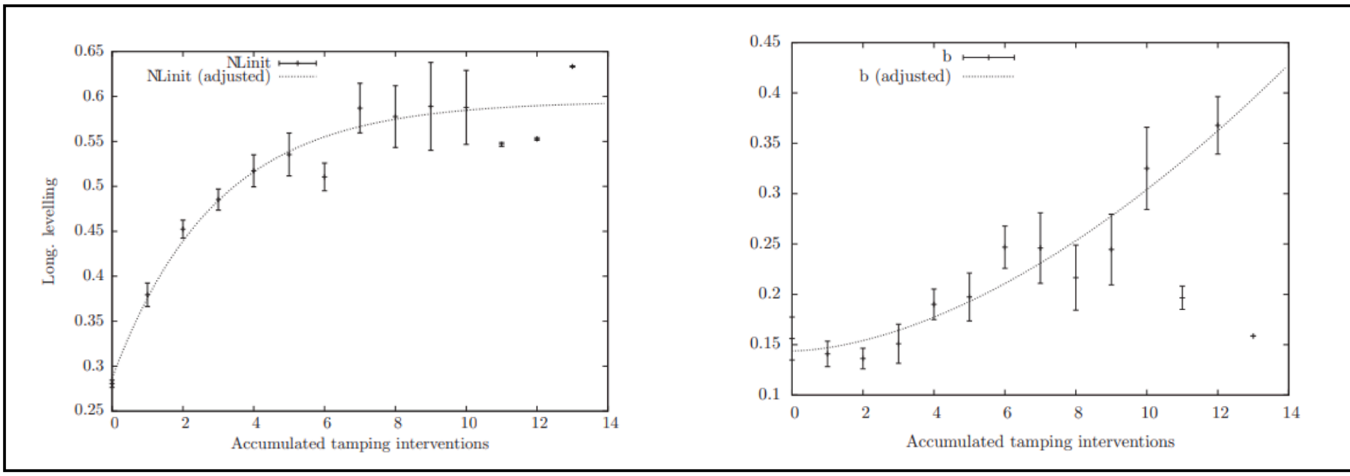

Fig. 3. *NLinit* (left) and b_n (right) after each tamping [25].

Fig. 4. Network Rail's New Measurement Train [27].

out every-six weeks, which provides a large history of the behaviour of the track. During an inspection, the train sensors (accelerometers) collect data in a raw format, which are immediately processed using the inertial method (double integration) for accelerometer data to obtain the final displacement readings. High-pass filters are applied to remove very long-wavelength forms (>100 m) and design features. For vertical profiles, transducers are used to filter out the displacement of the suspension system and provide a measurement of the actual rail surface.

3.2. Maintenance activities

Apart from geometry measurements, Network Rail has also provided access to maintenance records. These include high output tamping (also known as large-scale works for correcting multiple defects in a longer track span), local works, such as sprinter tamping and manual packing (against individual faults in short track spans), and welding and grinding. Welding and grinding activities are performed annually to remove imperfections on the rails, such as cracks and corrugation. These have not affected track geometry based on the available data and therefore are not currently part of the analysis. Large-scale works include the use of a dynamic track stabiliser (DTS) and is the most

Fig. 5. Network Rail tamping machine.

significant type of maintenance performed. Fig. 5 shows a tamping machine used by Network Rail.

As mentioned earlier, Network Rail employs two monitoring techniques as a means of scheduling interventions. The first one includes estimating standard deviation of geometry features for 200-metre track segments and studying them over time. Based on this value, maintenance activities may be required. An example of this is shown in Fig. 6. The large vertical drops are indicative of maintenance occurring at different points in time. The example shown includes both large-scale (red markers) and local (green markers) maintenance works. It can be seen that for this segment, over the lifespan of the track the standard deviation for vertical profile readings does not exceed 1.4 mm, which is indicating a satisfactory state of the railway. This type of behaviour dominates most of the line, as maintenance thresholds are fairly low on high-speed tracks, and it is constantly held to a very high standard of quality. Table 1 shows typical actions by Network Rail given different ranges of standard deviation values for vertical profile based on line speed. Tracks with a lower maximum allowable velocity present a lower risk for passengers, and therefore have a higher maintenance threshold than high-speed tracks.

Where TV, WV, AV, and SRV are the following:

- Target Value, where the reading is acceptable
- Warning Value which will require future maintenance
- Action Value, which is considered an intervention limit
- Speed Restriction Value, which is also an immediate action limit.

For the second monitoring technique, there are two exceedance levels when it comes to critical values of individual measurements – level 1 and level 2. Level 1 is set to identify a reading which is undesirable and will require attention in the future. This is not an immediate action fault as it does not pose a real danger to passengers. Level 2 defects, on the other hand, are considered to be more serious and must be rectified within specific timescales. Additional control measures are taken against cyclic top faults, which are repeated waveforms in close succession appearing on the rail profile, causing a higher track roughness. If the trace from the measurements shows that the surface of the rails is irregular and could lead to passenger discomfort, this may require tamping the section. Table 2 provides tolerance bands for individual measurements based on line speed. TSR in the Actions Required column stands for track speed restriction. All other abbreviations are as defined for Table 1.

The procedure for each maintenance activity performed by Network Rail is different and is therefore expected to have a varying effect on performance afterwards. High output tamping for example is typically performed over a longer portion of the track (several hundred metres or more), aiming to remove multiple defects within that span. In other

Table 1
Response actions based on 200-metre standard deviation values (via Network Rail).

Parameter 914	Vertical Alignment 35 m and 70 m SD Standard deviations are calculated for 200 m sections of track. (EN13848)			
Tolerance bands	High speed 201–300 km/h	Conventional Speed 81–200 km/h	Low Speed 80 km/h or less	Action Required
Good **TV**	< 1.0 mm	< 1.7 mm	< 3.8 mm	There are no SRV values as SD's are a guidance figure only and is used to highlight areas of concern rather than specific details. A WV should lead to a more thorough review of the 200 m section to highlight individual specific faults.
Satisfactory **WV**	1.0 – 1.5 mm	1.7 – 2.4 mm	3.9 – 5.4 mm	
Poor **AV**	> 1.5 mm	> 2.4 mm	> 5.4 mm	Inspect within 14 days Complete TEF 3016HS

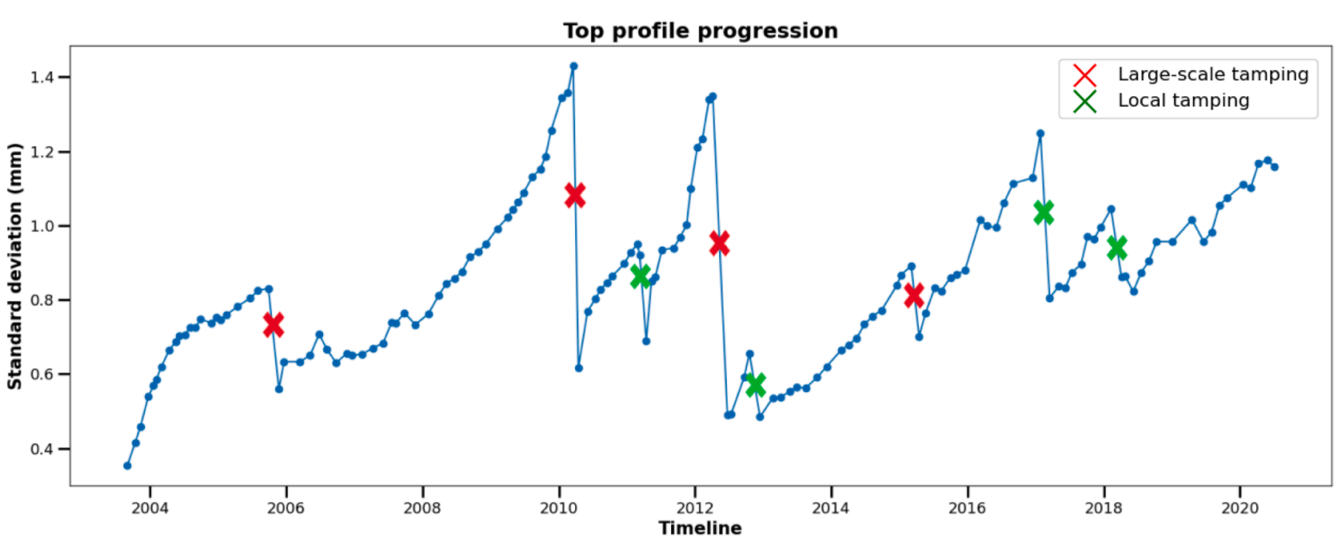

Fig. 6. Standard deviation values for a single 200-metre track segment (2003–2020) showing the quality of the left short-wave top profile (35 m) and maintenance works done during the period. Local tamping includes manual packing and sprinter tamping.

Table 2
Response actions based on individual measurements (via Network Rail).

Parameter 909	Vertical Alignment 35 m, RT35 or LT35			
Tolerance bands	High speed 201–300 km/h	Conventional Speed 81–200 km/h	Low Speed 80 km/h or less	Action Required
Construction	5 mm	+0, −30 mm	+0, −30 mm	New installation
Good TV	5 mm or less	9 mm or less	9 mm or less	None
Satisfactory WV	6 – 9 mm	14 – 17 mm	20 – 22 mm	Investigate root cause and plan for correction.
Poor AV	10 – 14 mm	18 – 25 mm	23 – 35 mm	Correct within 14 days
CRITCAL SRV	15 – 21 mm	26 – 27 mm	N/A	160 km/h TSR Inspect within 72 h Correct within 7 days
	22 – 25 mm	28 – 32 mm		80 km/h TSR
	26 – 37 mm	33 – 37 mm	36–37 mm	30 km/h TSR
	> 38 mm			Block the line

words, this reduces the track roughness. During this process, the track may be lifted if necessary. If that is the case, when the rails are lifted by a known amount, new ballast is placed under the sleepers, until the desired position of the track is obtained. This is known as a ballast drop. In this scenario, the old ballast is not removed, but simply packed using the tamper.

The other set of works (local tamping) are more focused and aim to eliminate individual defects at specific locations on the track. A sprinter tamper is used for this type of work. Once a fault has been identified by a recording vehicle, the ballast particle size and shape at that location are inspected. If they are worn out and rounded from the constant cyclic loads and past tamping works, then part of the material under the sleepers may be replaced at the location of the defect, after which it is packed. The new material under the sleepers is taken from the ballast shoulders, where it has not experienced traffic loading. The worn-out/spent ballast is then used to form the shoulder again.

If ballast is renewed, this is expected to stabilize the track condition at that location and potentially reduce the rate of deterioration, assuming this was the root cause of the defect. However, if this procedure is performed multiple times at one location, it may lead to worn out material being inserted under the sleeper again, which may not be effective. These activities are usually performed by Network Rail in September/October each year. This is different from high output tamping, where the old ballast is simply compacted under the sleeper using a tamper, which will cause further rounding and wearing out of the particles. Fig. 7 displays the difference in response of the standard deviation of the top left rail profile after both large-scale and local works within the same 50-metre track segment.

It may be noticed from Fig. 7 that the rates at which the standard deviation increases tend to be different based on the type of work performed. High output tamping appears to have increased the rate of deterioration immediately after the works, whereas following local corrections, the deterioration rate is generally not as rapid. This observation was analysed further, and the same type of behaviour is apparent in other parts of the studied line. This is suspected to be a result of introducing new ballast under the sleepers during sprinter tamping (local works).

Furthermore, shortly after the very first tamping of this section, which was performed in 2004, the maximum readings of the left profile increased, causing noticeable deterioration. Such a spike in readings is evident later on in 2019 as well (not shown above). This is a sign that while large-scale tamping may potentially correct one geometrical property, it could serve as a destabilizing event for another. This may especially be the case if the initial condition before tamping is not indicating faults. Therefore, it is important to execute targeted maintenance to remove specific defects and not worsen others. Results show that on average, deterioration rates after large-scale works are approximately 42 % higher than after sprinter tamping for vertical profiles and 63 % higher for twist measurements.

4. Methodology

4.1. *Artificial Neural Network (ANN)*

The work in this study employed a single-hidden-layer artificial neural network (ANN). This is also known as a multilayer perceptron. It is a feed-forward supervised learning algorithm, which requires a labelled data set during training. Fig. 8 shows the architecture of such a model. Each neuron has an associated weight (w_{ji}^{k}) which connects it to the neurons of the previous layer and each layer also has bias values (b_{j}^{k}), which can regulate when a neuron becomes meaningfully active. The data to be processed is of the form $\mathbf{X} = \{x_1, x_2, \cdots, x_n\}$ and is passed through the input layer to each neuron of the first hidden layer as a weighted sum of these values, based on its associated weights. This sum

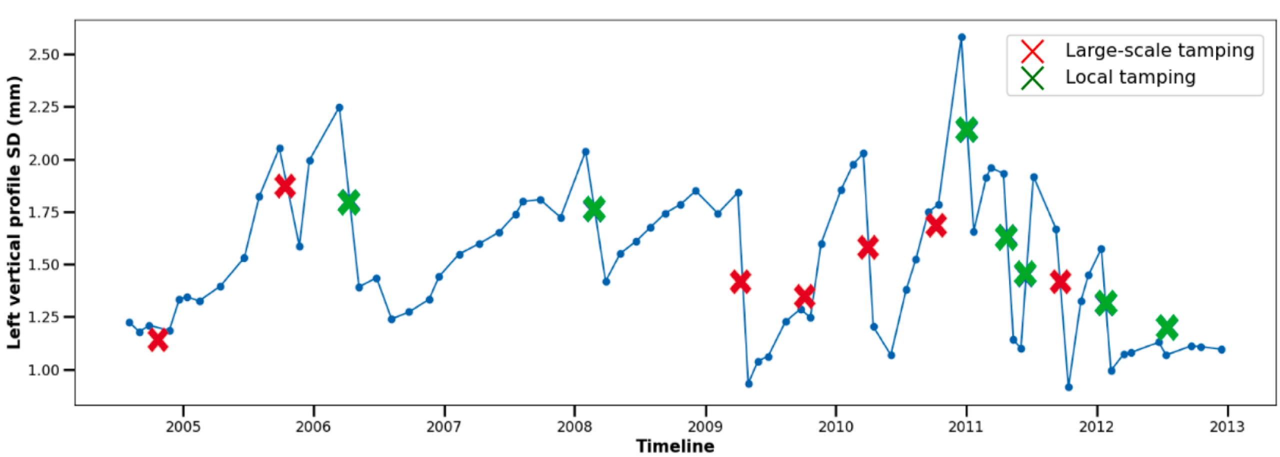

Fig. 7. Instances of track response after large-scale and local tamping.

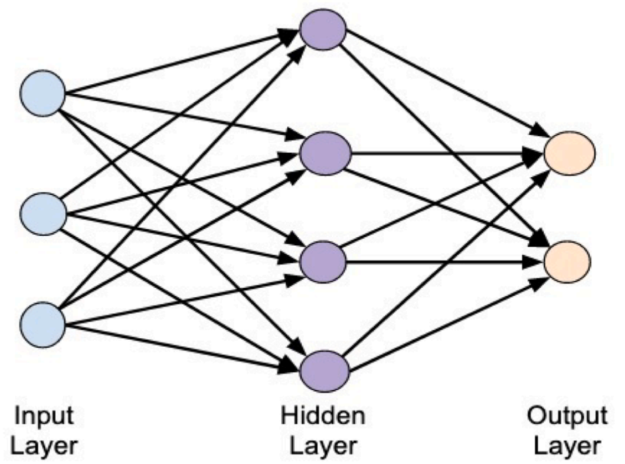

Fig. 8. Single hidden layer ANN model schematic.

is then processed using an activation function before the output is passed onto the next layer. This is shown in the equation below.

$$a_j^1 = f\left(\sum_i w_{ji}^1 \times x_i + b_j^1\right) \tag{3}$$

Here a_j^1 is the output value of the jth neuron in the 1st hidden layer of the network; $f(x)$ is the layer activation function; the argument in the function is the value passed on to the hidden layer from the previous one. w_{ji}^1 is the weight connecting the jth neuron of the first hidden layer to the ith neuron of the input layer, x_i is the value in the ith neuron of the input layer, and b_j^1 is the bias of the jth neuron in the hidden layer. This process is then repeated, and the new weighted sum of the current layer is passed onto the next hidden layer (or the output layer if there is only one hidden layer in the model), where a new activation function is applied. This step can be represented using a more general expression for the entire layer:

$$a_j^k = f\left(W_{ji}^k \times a_i^{k-1} + b_j^k\right) = f(z_j) \tag{4}$$

Here, $a_j^k = \{a_0, a_1, \ldots, a_n\}$ is a vector containing neuron values from the kth layer (similarly for a_i^{k-1}), W_{ji}^k and b_j^k are a matrix and a vector containing the weights and biases for the kth layer in the network, and z_j is a simplified version of the argument inside the activation function.

There is a variety of activation functions available, some of the more popular ones are sigmoid functions, hyperbolic tangent, rectified linear unit (ReLU), among others (Fig. 9). The suitability of each one is often judged based on their differential properties. When the data reaches the output layer, a single value (or a vector) is produced which is indicative of the label of a given example (in the case of classifiers). During the training process, a set of training examples is passed through the network, the predicted labels are compared to the expected ones and the difference between the two is summarised using a loss function (L). The further away the predicted values are from the desired ones, the larger L will be. Similarly to activation functions, there is a variety of loss functions as well, which are well-suited for different problems. Mean Squared Error (MSE) for example is a popular loss function to use mainly in regression problems, although it is not appropriate for handling outliers in the data, since it squares the errors. Mean Absolute Error (MAE) on the other hand, is more robust when dealing with outliers and can also more effectively account for negative errors. There are a number of other functions available, each of which may be more suitable for a specific problem and data set.

Initially, the model will not perform well on training data set, at which point a backpropagation process begins to minimize this loss function [28]. This is done using the stochastic gradient descent method [29]. Each time the loss function is calculated for a given set of examples, gradient descent works its way back to the input layer and corrects the trainable parameters (weights and biases) to reduce the loss. The passage of the entire training set in smaller batches of examples (or as a whole) is known as one epoch. Minimization works by estimating the sensitivity of the loss function with respect to the weights and biases at each layer, i.e., the gradient. Selecting the optimal activation and loss functions for the specific problem helps the convergence. To estimate the gradient, the chain rule is applied.

$$\frac{\partial L_0}{\partial W_{ji}^k} = \frac{\partial z_j^k}{\partial W_{ji}^k} \times \frac{\partial a_j^k}{\partial z_j^k} \times \frac{\partial L_0}{\partial a_j^k} \tag{5}$$

where L_0 is the loss calculated from the first training example. The total loss is the average of all training examples.

$$\frac{\partial L}{\partial W^k} = \frac{1}{n}\sum_{l=0}^{n-1} \frac{\partial L_l}{\partial W^k} \tag{6}$$

This is the gradient of the loss only with respect to the weights at the kth layer. The same equation is applied with respect to the biases as well. The gradients with respect to weights and biases at each layer comprise the gradient vector $\nabla L = \left\{\frac{\partial L}{\partial W^1}, \frac{\partial L}{\partial b^1}, \cdots, \frac{\partial L}{\partial W^n}, \frac{\partial L}{\partial b^n}\right\}$. Each component of the gradient vector holds information about the sensitivity of the loss function with respect to the weights and biases at each layer. This process is repeated until L is sufficiently low, indicating that the model has

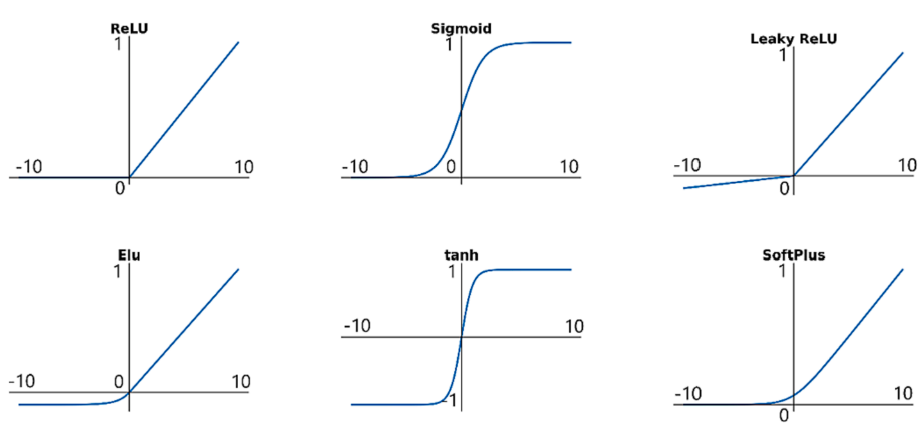

Fig. 9. Different activation functions.

learned crucial patterns in the data set and is able to predict labels at a high rate. Further details about the development and applications of Artificial Neural Networks can be found in [30].

5. Data processing and initial model implementation

5.1. Track classification

A maintenance activity will not always result in the expected theoretical improvement in quality, since it is an imperfect event which can be affected by a number of factors. One of the goals of this study was to estimate an approximate rate of efficiency of tamping works i.e., how often an activity has made a noticeable improvement. Quality improvement here is defined as a reading at a unique location decreasing by at least 1–2 mm in between consecutive inspections. Variations smaller than that are difficult to justify due to measurement accuracy. Similarly, an increase in readings of>1 mm was considered deterioration. The classification criterion selected is relative to the available data set. The threshold could be set higher for a different track which exhibits more significant changes in readings. However, the high-speed line analysed here is of very high quality and rapid settlement between inspections are extremely rare, making it implausible to train a model to recognize only larger changes.

To perform an initial classification, the track was broken down into small segments (50 m long) and a number of features were extracted from the associated measurements. A single-hidden layer Artificial Neural Network (ANN) was employed using these features to classify each segment at a given point in time as “deteriorated” – class 1, “improved” – class 2, or “not changed” – class 3.

5.2. Spatial alignment

Firstly, readings from the two inspections must be aligned spatially, so the peaks and troughs of the two signals match and allow for a comparison to be made. To do this, cross-correlation between the signals was used – this takes one of the signals and slides it across the other one, until a good match is found. This process is also known as a sliding dot product, which can indicate by how many steps the first set of measurements is lagging behind the second and the two can be aligned. Fig. 10 displays the results from the operation. This process has been made automatic for all measurements.

5.3. Estimation of deterioration rate

Once individual inspection readings have been aligned, the next step is the estimation of a rate of deterioration of track geometry. The following equation was developed:

$$Deterioration\ rate = \frac{(|reading_1| - |reading_2|)*30}{Time\ between\ inspections\ [days]}\left[\frac{mm}{month}\right] \quad (7)$$

Readings 1 and 2 are from two separate (consecutive) inspections. This is done for all individual readings from a segment. The difference between the two is divided by the number of days between inspections and multiplied by 30 to obtain an approximate deterioration rate in units of mm/month. The absolute values of the measurements have been used, as the difference between them will show if a reading is closer to or further away from an ideal track condition (zero measurement). This idea applies to both positive and negative geometry readings. A negative rate will indicate deterioration, whereas a positive value is showing an improvement. Fig. 11 displays this – the top half of the figure is showing consecutive inspections, where the one in April 2014 (orange line) is showing improvement between 250 m and 300 m. This results in an increase in the rate of deterioration in the plot below it, which can be automatically detected. The bottom half of the figure is showing the same phenomenon, although for deterioration. Each segment on the track will produce a different set of rates, which are used to help classify its condition. The number of values in each set will be defined by the number of measurements in the segment (typically 250). Two features related to these sets are used as input to the neural network algorithm – the standard deviation and skewness.

5.4. Other input features

The algorithm requires other features as well for optimal performance, the main of which is the ratio of the standard deviations of each inspection (SD_1/SD_2). If this ratio is very close to 1 (typically between 0.9 and 1.1), that means that little change has occurred since the most recent inspection, whereas a value much different from 1 indicates that there has been some change in the readings. Fig. 12 shows how this ratio varies along with the SD of the deterioration rates in the same segment at different points in time. These are the two key input features necessary to detect a pattern in the data.

5.5. Model training and performance

The network consists of 3 layers in total – an input layer, hidden, and

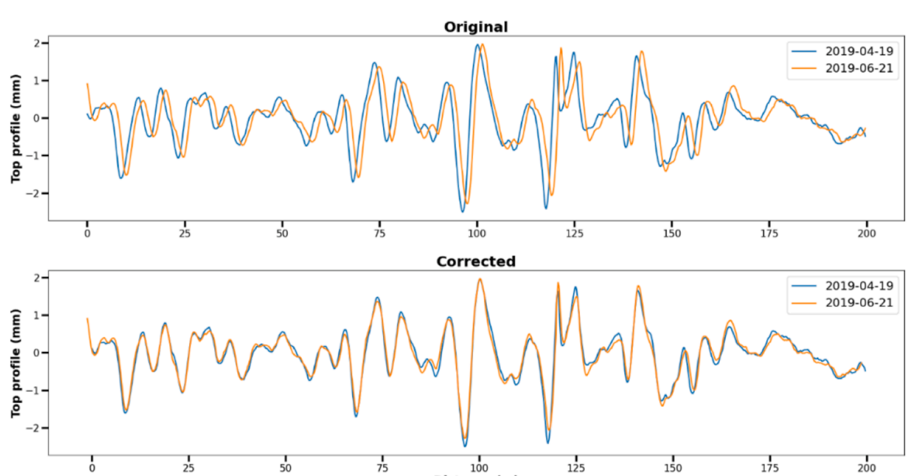

Fig. 10. Signal alignment using cross-correlation.

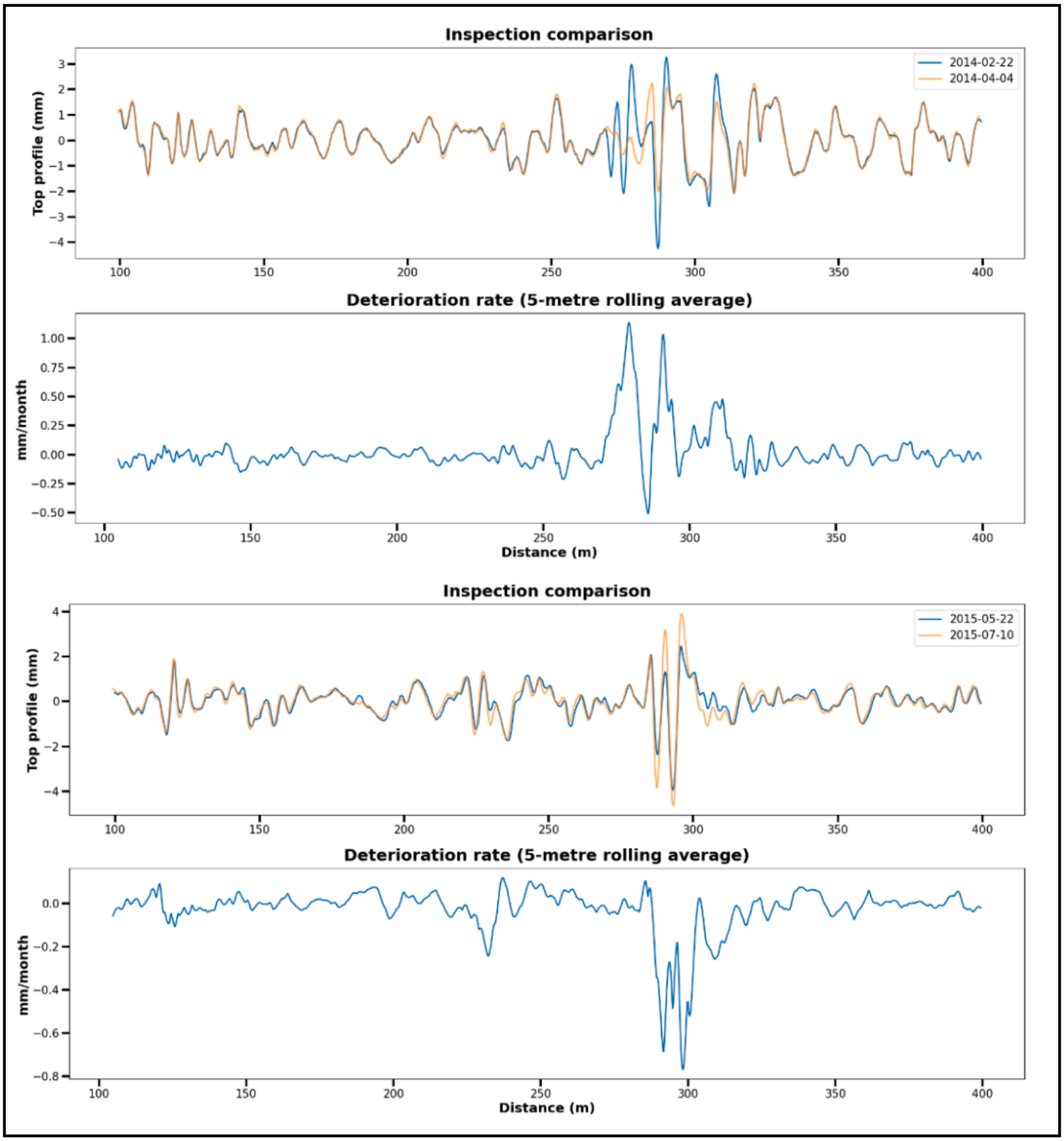

Fig. 11. Deterioration rates indicating improvement (top) and worsening (bottom) in quality. The blue and orange lines in the inspection comparison plots show the two consecutive inspections and the regions which have experienced a change. The deterioration plots show a rolling average for a smoother and more interpretable line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

an output layer (see Fig. 8). The number of neurons in each layer were varied, but it was found that the model performed best when the input layer consisted of 50 neurons, the hidden layer of 15, and 3 neurons in the output layer (one for each class). The initial weights of the model were generated using a random normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{8}$$

It was found that a mean absolute error (MAE) loss function performed the best. The three classes indicating segment behaviour (improved, deteriorated, no change) were encoded, a ReLU activation function (equation (10)) was used in the hidden layer and a SoftMax function (equation (11)) was used for the output layer:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \widehat{Y}_i| \tag{9}$$

$$f_i(\overrightarrow{x}) = max(0, x) \tag{10}$$

$$f_i(\overrightarrow{x}) = \frac{e^{x_i}}{\sum_j e^{a_j}} \tag{11}$$

The geometric feature used to train the network described here was top profile (35 m), but the same technique has shown to be applicable to other parameters as well (twist and long-wave top profile). The training set consisted of approximately 4000 examples, but was later reduced to ~ 2500, due to an imbalance in the data. The majority of the examples were in class 3 (approximately 90 %), due to the high quality of the track. This did not allow the model to train adequately [31], and therefore some of these examples were removed. 30 % of the data was used for testing purposes during training, and the rest was used as training data. A further 25 % of the training data was used for validation purposes.

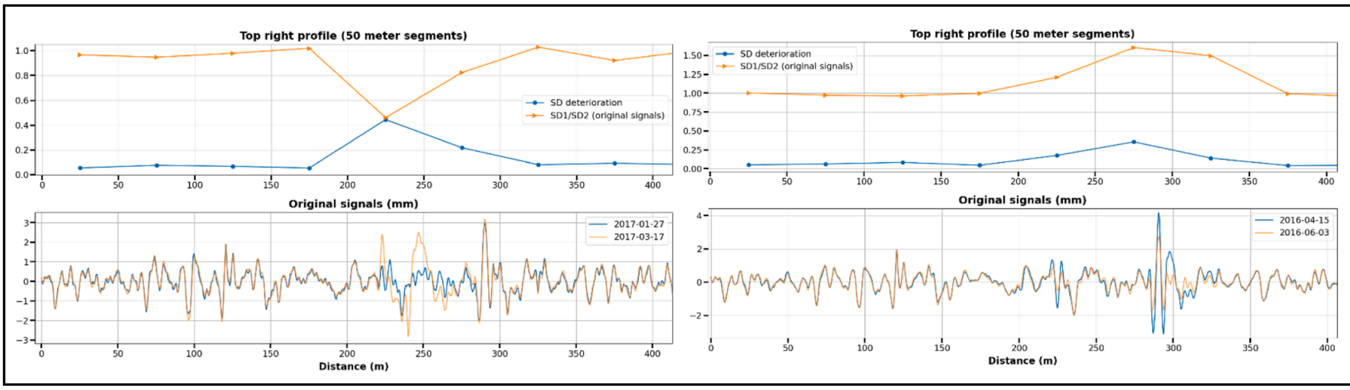

Fig. 12. Variation of deterioration rate SD and standard deviations ratio (SD_1/SD_2) in a worsened portion of the track (left) and in an improved portion of the track (right).

Fig. 13 shows the learning curve (accuracy) of the best performing model. It achieved an accuracy of approximately 98 % on the validation set during training. Upon testing on a new unseen data set, the model exhibited a high sensitivity when classifying deterioration. This required further output processing, where segments with a SD_1/SD_2 ratio very close to 1 were manually changed to class 3. It is expected that given a different data set which shows more variation in measurements, the model will be able to automatically make this distinction, although this was not available at the time of the study.

Following post-processing, the overall accuracy of the model on an unseen data set was 99.6 %, although the data used for this evaluation was still highly imbalanced. The majority of the examples were class 3, which the network does not struggle to identify. The detection accuracy for class 2 (maintenance) on the track was 88.9 % and for class 1 (development of defects) – 77.5 %. It is suspected that the better performance for detecting class 2 is partly due to the fact that maintenance results in a large absolute change in readings. On the other hand, deterioration is much more gradual and therefore not as statistically significant, making class 1 more difficult to detect. Upon closer examination, it was found that the weaker performance in identifying deterioration is mainly due to small variations which are very close to the decision boundary of the model. Fig. 14 provides examples of the classification abilities of the developed model.

Such form of modelling work shows the potential of machine learning algorithms to process large amounts of track geometry data in a short time frame and their ability to detect anomalies in track behaviour. By performing this analysis, regions of the track experiencing settlement and requiring maintenance may be more accurately identified in order to remediation works where they are needed. Reducing the amount of healthy track being tamped will ensure preservation of ballast and the extension of the asset's lifespan.

5.6. Maintenance efficiency

Since the described algorithm is able to detect fault corrections on the track reasonably well (88.9 %), it was cross-referenced with records of maintenance in order to estimate how much of the tamped track had actually improved. This was done for the primary types of maintenance – high output tamping, sprinter tamping/manual packing, and tamping on switches and crossings (S&C). S&C tamping is considered on its own since the track in those regions is unique and the effects may be different. Both short-wave (35 m vertical profiles and twist) and long-wave (70 m mean vertical profile) features were analysed. The percentage of track positively influenced after maintenance is shown in Table 3.

The top vertical profiles (short-wave) and the twist measurements

Fig. 13. Model accuracy during training.

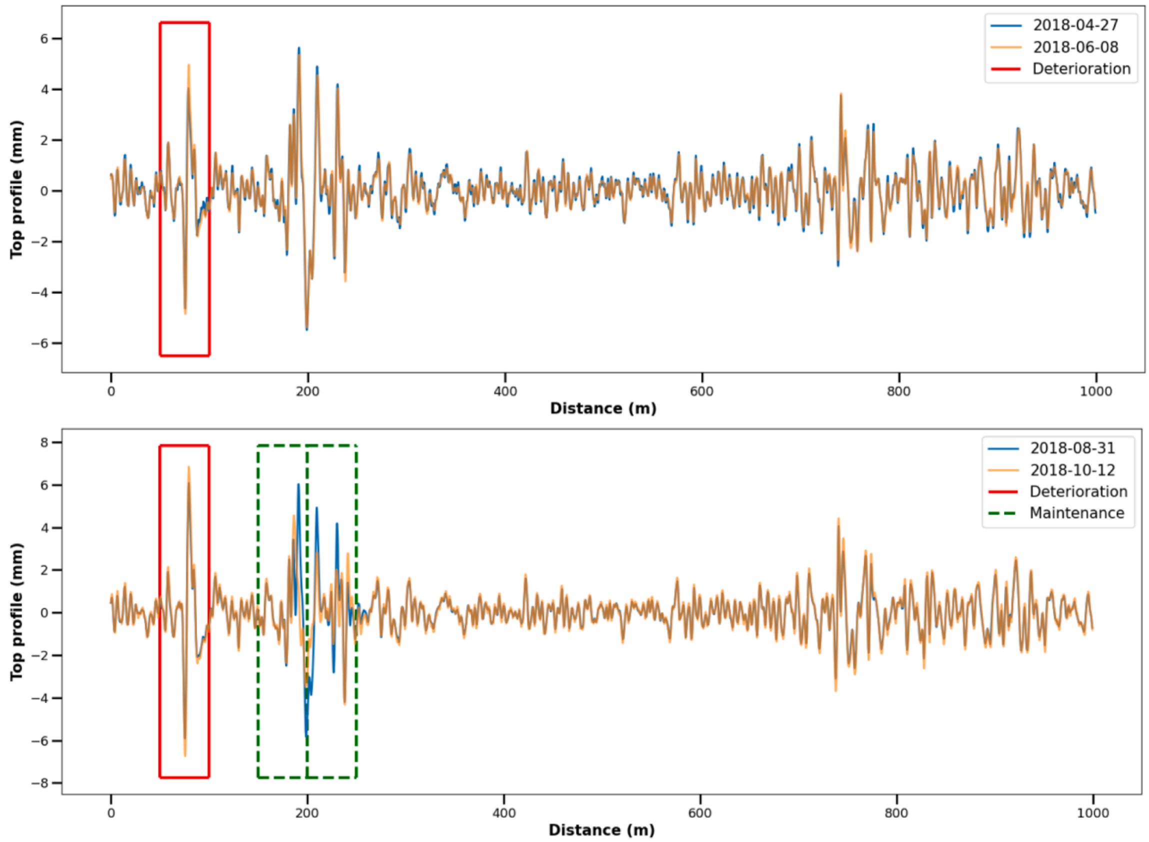

Fig. 14. Model comparison versus original measurements.

Table 3
Estimated approximate maintenance efficiency.

Activity	High output tamping	Sprinter tamping/ Manual packing	S&C tamping
Track segments with improved short-wave vertical profile	30.1 %	78.1 %	34.5 %
Track segments with improved twist (3 m)	36.7 %	67 %	28.6 %
Overall amount of track segments with improved short-wave features	45.8 %	84.9 %	39.8 %
Track with improved long wave profile faults	36.9 %	55.7 %	35.8 %
Combined long- and short-wave features	52.5 %	84 %	45.6 %

have been examined separately, as well as together. Similarly, long-wave profile was also checked individually. All features were then combined to come up with an overall efficiency value. Not all of the tamping records indicate the exact locations of tamping, but a specific asset where several smaller segments have been tamped. The impact of this has been removed as much as possible, however it may still have a residual effect and reduce these values (mainly for high output tamping). In addition, while most of the maintenance activities are known, there are also some missing records of both high output tamping and sprinter tamping/manual packing. If the full records are available, the efficiency may be slightly higher.

A total of 858 50-metre segments were analysed for all specified geometrical features. Measurements cover a 10-year period (2007–2016) when maintenance records are available. Only a small portion of the segments which improved were in critical condition prior to the maintenance. In most cases, the associated readings were just large enough as to exhibit a noticeable change (>1mm) following corrections. As expected, high output tamping, which covers longer sections of track, has a lower efficiency than sprinter tamping across all features. This is due to the fact that when a longer segment of a track is tamped, the machine also covers areas where the ballast is in good condition. In that case the changes in readings are either not noticeable, or they may possibly become worse. The latter is a rare event, although it may still occur, especially after large-scale works. In some instances when the track appears to have deteriorated after maintenance, this may be due to a very quick recovery in readings shortly after a correction.

6. Conclusions and future work

This paper has explored the history of a high-speed railway line in the UK and investigated ways of automatically detecting sudden changes in track geometry measurements. The aim is to help assess the current state of the railway, as well as the efficiency of maintenance works. A large data set of track geometry measurements was analysed:

1. An Artificial Neural Network was trained on the obtained data set to identify track segments exhibiting a change in quality in the period between inspections.
2. The track was divided into three possible classes – showing an improved state (maintenance), showing a more deteriorated state since the last inspection, and showing no change.
3. The first two of the above classes achieved an accuracy of 88.9 % and 77.5 %, respectively.
4. The model was used to estimate how often specific types of maintenance activities have had a positive effect on a given segment.

5. It was found that high output tamping, which is undertaken over longer sections of track, exhibits a lower rate of efficiency than local tamping works which are used for removing individual defects. The explanation for this is that when tamping is undertaken on a large scale, the machine will also tamp track of high quality, which will show no noticeable improvements, or possibly deterioration.
6. Historically, the decision of whether to tamp a track is largely based on the progression of the standard deviation of a certain geometric feature from a 200-metre track segment, or another similar index. However, it is suggested here that these measures may be sensitive to the development of individual defects and inaccurately indicate faults in the entire segment. In this study, a shorter segment length of 50 m was used aiming to better localise the development of faults.
7. Using the analysis techniques shown in this paper one will be able to plan more cost effective and targeted tamping campaigns. It may prove more cost effective to use sprinter tamping and local hand packing at specific selected sites. Where long sections of track are deteriorated, high output tampers would be more appropriate.

Since track faults appear to be mainly affecting local regions, it will be beneficial to detect their more exact location. Future research will include the optimization of similar models aimed at monitoring the development of such faults more closely, as well and identifying specific reasons for deterioration. In doing so, root-causes of faults can be removed and track quality stabilised in the long-term.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

This is part of a four-year PhD project at the University of Edinburgh, jointly funded by the Scottish Research Partnership in engineering (SRPe) – SRPe-IDP - and Loram UK (industry) – CT-7381. Data has been provided by Network Rail.

References

[1] Dataportal.orr.gov.uk. 2022. Passenger rail usage | ORR Data Portal. [online] Available at: <https://dataportal.orr.gov.uk/statistics/usage/passenger-rail-usage> [Accessed 18 May 2022].
[2] GOV.UK. 2020. UK Becomes First Major Economy To Pass Net Zero Emissions Law. [online] Available at: <https://www.gov.uk/government/news/uk-becomes-first-major-economy-to-pass-net-zero-emissions-law.> [Accessed 17 November 2020].
[3] High Speed 2. 2020. Carbon - Putting Britain On Track To A Net Zero Carbon Future | High Speed 2. [online] Available at: <https://www.hs2.org.uk/why/carbon/> [Accessed 17 November 2020].
[4] T. Lidén, Railway Infrastructure Maintenance - A Survey of Planning Problems and Conducted Research, in: Transportation Research Procedia, Elsevier, 2015, pp. 574–583.
[5] A. Falamarzi, S. Moridpour, M. Nazem, A review of rail track degradation prediction models, Aust. J. Civ. Eng. 17 (2) (2019) 152–166.
[6] C. Charoenwong, D.P. Connolly, P.K. Woodward, P. Galvín, P. Alves Costa, Analytical forecasting of long-term railway track settlement, Comput. Geotech. 143 (2022) 104601.
[7] I. Grossoni, W. Powrie, A. Zervos, Y. Bezin, L. Le Pen, Modelling railway ballasted track settlement in vehicle-track interaction analysis, Transp. Geotech. 26 (2021) 100433.
[8] J. Sadeghi, Development of Railway Track Geometry Indexes Based on Statistical Distribution of Geometry Data, J. Transp. Eng. 136 (8) (2010) 693–700.
[9] H. Khajehei, A. Ahmadi, I. Soleimanmeigouni, A. Nissen, Allocation of effective maintenance limit for railway track geometry, Struct. Infrastruct. Eng. 15 (12) (2019) 1597–1612.
[10] C. Vale, M.S. Lurdes, Stochastic model for the geometrical rail track degradation process in the Portuguese railway Northern Line, Reliab. Eng. Syst. Saf. 116 (2013) 91–98.
[11] J. Andrews, D. Prescott, F. De Rozières, A stochastic model for railway track asset management, Reliab. Eng. Syst. Saf. 130 (2014) 76–84.
[12] A. Andrade, P. Teixeira, Hierarchical Bayesian modelling of rail track geometry degradation, Proc. Inst. Mech. Eng., Part F: J. Rail Rapid Transit 227 (4) (2013) 364–375.
[13] K. Popov. Conversation with Matthew Rippin, David Whitehead of Network Rail. 1 March 2021.
[14] K.Y. Koo, J.M.W. Brownjohn, D.I. List, R. Cole, Structural health monitoring of the Tamar suspension bridge, Struct. Control Health Monit. 20 (4) (2013) 609–625.
[15] M. Fraser, A. Elgamal, X. He, J.P. Conte, Sensor network for structural health monitoring of a highway bridge, J. Comput. Civ. Eng. 24 (1) (2010) 11–24.
[16] A. Guo, A. Jiang, J. Lin, X. Li, Data mining algorithms for bridge health monitoring: Kohonen clustering and LSTM prediction approaches, J. Supercomput. 76 (2) (2020) 932–947.
[17] H. Guler, S. Jovanovic, G. Evren, Modelling railway track geometry deterioration, Proc. Inst. Civ. Eng. – Transp. 164 (2) (2011) 65–75.
[18] C. Meier-Hirmer, A. Senée, G. Riboulet, F. Sourget, M. Roussignol, A decision support system for track maintenance, Comput. Railways X (2006).
[19] S. Costello, A. Premathilaka, R. Dunn, Stochastic Rail Wear Model for Railroad Tracks, Transp. Res. Rec.: J. Transp. Res. Board 2289 (1) (2012) 103–110.
[20] P. Xu, R. Liu, F. Wang, F. Wang, Q. Sun, Railroad Track Deterioration Characteristics Based Track Measurement Data Mining, Mathematical Probl. Eng. 2013 (2013) 1–7.
[21] A. Falamarzi, S. Moridpour, M. Nazem, S. Cheraghi, Development of a random forests regression model to predict track degradation index: Melbourne case study, Australian Transport Research Forum (ATRF), Darwin, 2018.
[22] C. Hu, X. Liu, 'Modelling Track Geometry Degradation Using Support Vector Machine Technique', in: Proceedings of the 2016 Joint Rail Conference. April 12-15, 2016, Columbia, SC, USA (2016).
[23] J. Sadeghi, H. Askarinejad, Application of neural networks in evaluation of railway track quality condition, J. Mech. Sci. Technol. 26 (1) (2012) 113–122.
[24] L. Quiroga, E. Schnieder, A heuristic approach to railway track maintenance scheduling, Comput. Railways XII (2010).
[25] L.M. Quiroga, E. Schnieder, Monte Carlo simulation of railway track geometry deterioration and restoration, Proc. Inst. Mech. Eng., Part O: J. Risk Reliability 226 (3) (2012) 274–282.
[26] M. Sasidharan, M.P.N. Burrow, G.S. Ghataora, A whole life cycle approach under uncertainty for economically justifiable ballasted railway track maintenance, Res. Transp. Econ. 80 (2020) 100815.
[27] Network Rail, 2020, New Measurement Train, https://www.networkrail.co.uk/running-the-railway/looking-after-the-railway/our-fleet-machines-and-vehicles/new-measurement-train-nmt/.
[28] M. Buscema, Back Propagation Neural Networks, Subst. Use Misuse 33 (2) (1998) 233–270.
[29] S. Ruder. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1600.04747, 2017.
[30] Y.C. Wu, J.W. Feng, Development and application of artificial neural network, Wireless Pers. Commun. 102 (2) (2018) 1645–1656.
[31] F. Thabtah, S. Hammoud, F. Kamalov, A. Gonsalves, Data imbalance in classification: Experimental evaluation, Inf. Sci. 513 (2020) 429–441.