Systems thinking approach for improving maintenance management of discrete rail assets: a review and future perspectives

Yue Shang, Maria Nogal, Haoyu Wang & A. R. M. (Rogier) Wolfert

To cite this article: Yue Shang, Maria Nogal, Haoyu Wang & A. R. M. (Rogier) Wolfert (2021): Systems thinking approach for improving maintenance management of discrete rail assets: a review and future perspectives, Structure and Infrastructure Engineering, DOI: 10.1080/15732479.2021.1936569

To link to this article: https://doi.org/10.1080/15732479.2021.1936569
Systems thinking approach for improving maintenance management of discrete rail assets: a review and future perspectives

Yue Shanga, Maria Nogalb, Haoyu Wanga, and A. R. M. (Rogier) Wolferta

Engineering Asset Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands; Railway Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands

ABSTRACT
Performance evaluation and maintenance planning are gaining importance with ageing rail infrastructure and increasing demand on track safety and continuous availability. The discrete/point railway assets (e.g. bridges, level crossings) together with extended track sections constitute the main railway network infrastructure. The former has important implications in train safety, riding comfort and operating expenditures due to local intensified degradation and plays a role in effective network capacity due to their large quantity. The heterogeneity in asset features and operating environment also adds difficulties to efficient maintenance planning of multiple discrete assets. The current review screens the issue to level crossings, as little concern has been engaged to this asset type, and draws together different perspectives related to their maintenance management. The systems thinking approach is integrated and two levels of asset management (i.e. micro- and macro-level) are used to structure the synthesis, which are interdependent and synergistic. Two major approaches, namely, the mechanistic and data-driven modelling are synthesised. Both contribute to the maintenance knowledge and their comparisons are elaborated. Limitations in existing studies are identified and directions for future research are provided, aiming to contribute to a more refined ‘inspection and diagnosis’ process to properly capture the local track issues and move towards system-level maintenance approach for multiple level crossings.

1. Introduction

Maintenance interventions are necessary for proper functioning and lifetime extension of rail infrastructure but are costly considering long lifespans of the infrastructure. Conservative maintenance may lead to unnecessary activities and cost overruns, while late interventions imply safety hazards to the train operation and service disruptions. Therefore, planning for maintenance can be uncertain and complex, especially when considering the network infrastructure as the assets within a network often exhibit heterogeneity in constituting components, operational regime and environmental condition.

The network infrastructure consists of extended track sections and supporting assets such as bridges and level crossings. These supporting assets are geographically distributed and limited in length, which can be conceptualised as point assets connecting the open track to support the network functioning. Since the railway track is a linear and non-redundant system, the track safety and availability depend directly on the condition of the individual track sections and point assets (Khajehei, Haddadzade, Ahmadi, Soleimanmeigouni, & Nissen, 2020). Unlike the open track, the construction form of the point assets is not generally consistent.

For example, in the level crossings, the ballasted track is partly stiffened by rubber/concrete panels to form a road surface (Le Pen et al., 2014). Alternatively, a slab track section is formed at the crossing area, e.g. embedded rail system, as shown in Figure 1. Prefabricated concrete slabs are placed on compacted soil layers and naturally form a road surface with the adjacent pavement. The rails are embedded in concrete grooves and fixated by an elastic compound (Matias & Ferreira, 2020). The former case represents a change in track superstructure, and the latter forms a transition from the normal embankment (where a ballasted track places) onto a hard substructure (the concrete slab) and superstructure components also differ (e.g. fastening).

Due to the sudden change in the structural solution and/or geotechnical foundations, the approaches to the point assets are prone to strong amplification of the responses and localised degradation. The areas are commonly designated as track longitudinal transitions (referred as track/rail transition, hereon) characterised by the presence of (abrupt) longitudinal variations of vertical stiffness (referred as stiffness (longitudinal) variation, hereon) and differential settlement (track geometry irregularity).

The track geometry degradation is mostly used to characterise the overall track performance and considered the
primary degradation mechanism in train-track systems. It has important implications in safety, riding quality, and maintenance costs of the overall infrastructure. For the point assets, as an example shown in Figure 2, four longitudinal level measurements are recorded at a level crossing on the Dutch railway network: the red dotted lines indicate the junctions of the crossing and approaches, where both sides are susceptible to differential settlement and they show different patterns influenced by the moving direction. This geometry problem amplifies wheel-rail forces that induce additional causes for local differential settlement and loss of contact under sleepers, further accelerating the local degradation.

The point assets represent a large asset group in a network. For example, the British railway has a historical stock of level crossings over 6500 running on 15,847 kilometres of the track (ORR, 2019). The Dutch national railway network has a total mileage over 7000 kilometres including 2589 level crossings and 455 railway bridges (ProRail, 2019). The point assets are not extended in length like the open track, but they represent critical track areas susceptible to local intensified degradation. They generally vary in length, configuration, and constituting properties; operational and environmental conditions may also differ in place. The heterogeneity results in varied asset degradation behaviour and infrastructure managers often struggle to ascertain the actual condition in a large point asset population and decide the optimal timing of interventions to preserving the deteriorating assets over a network.

There are several reviews concerning track geometry degradation and maintenance modelling (Higgins & Liu, 2017; Soleimanneigouni, Ahmadi, & Kumar, 2016), big data analytics for rail track maintenance (Ghofrani, He, Goverde, & Liu, 2018; Nakhaee, Hiemstra, Stoelinga, & van Noort, 2019), mathematical algorithms for maintenance planning and scheduling problems (Lidén, 2014). These reviews focus on defect treatment of the open track, while the point assets, problematic areas with recurrent geometrical issues and other degradation features are not distinguished. Turnouts are critical groups among the point assets and a detailed review on their degradation modelling has been provided; however, maintenance-related modelling techniques are not included (Minbashi, Bagheri, Golroo, Khouy, & Ahmadi, 2016). The reviews regarding the railway bridges are also available, which mostly concern the track-bridge dynamics (Arvidsson & Karoumi, 2014; Zhai, Han, Chen, Ling, & Zhu, 2019).

Considering the large quantity, scattered locations, localised degradation features, and potential heterogeneity of the point assets, the question is how best to apply maintenance to keep the condition of a group discrete rail assets at an acceptable level, in terms of both safety and service continuity, in a cost-efficient manner. The current paper aims at drawing together different perspectives related to their maintenance management, with particular emphasis on the level crossings, as this asset type forms a large stock in the rail asset groups but has not received much attention in the literature. As a jointly used area by railway and highway traffic, its functioning has important implications in safety and operating performance of both traffic. By synthesising the relevant studies, this paper aims to discuss omissions and limitations in the previous research and propose avenues for efficient maintenance management of multiple level crossings, which also sheds light on the other discrete rail assets.

The asset condition is featured by temporal evolution and spatial variation. The track safety, cost, possession and service continuity are common factors influencing the maintenance decisions. To map out the maintenance design, this paper integrates systems thinking into the conception of maintenance and two levels of asset management are recognised, where the decision is manifested at the micro-level and macro-level. The micro-level often involves design features of a single asset. By establishing trackside experiments and/or formulating models that characterise track configuration and constituting properties, localised track behaviour under train loading and the corresponding preservation techniques can be evaluated. The macro-level analysis takes a global view on multiple physically related assets in a network, aiming to track their physical condition over time by inspection/monitoring and determine optimal timing of maintenance based on degradation modelling.

Through a bottom-up mechanism, the two levels of asset management are meant to be interdependent and synergistic. The synthesis of the relevant studies is therefore organised by following the bottom-up approach: Section 2 focuses on the micro-level analysis and discusses field experiment cases and modelling techniques. The existing mechanistic models that analyse the track behaviour at transitions to the point assets are synthesised and classified into transient dynamic analysis and long-term numerical prediction, in sub-sections 2.2 and 2.3, respectively. Section 3 explores the tasks involved in macro-level maintenance management, which builds up from track condition measurement and characterisation (3.1), degradation modelling (3.2) to maintenance planning (3.3). Section 4 discusses omissions and limitations in existing studies and proposes avenues for future research. Section 5 ends with concluding remarks.

2. Micro-level experimental investigation and mechanistic modelling

The micro-level analysis involves establishing field monitoring to gain insight into the track behaviour and developing
mechanistic models to simulate track dynamic responses during train passage. There are a few studies dedicated to field instrumentation on the level crossings. The behaviour of a level crossing before and after renewal has been evaluated in the UK (Le Pen et al., 2014). Geophones and digital image correlation were instrumented at the approach to measure sleeper deflections. It was diagnosed that hanging sleepers evident before tamping reoccurred very soon afterwards, indicating that tamping was not effective in removing hanging sleepers on the studied approach.

The case in the US is different that crossing areas, under the combined highway and railway traffic, settle more than the adjacent approaches and gradually become a low spot in the vertical profile (Rose, 2011). The reason is that the rail crossings provide a lower load-bearing capacity due to the flexible all-granular trackbed and are prone to excessive deflections, causing rapid abrasion and wear of the crossing components. The surface water penetrates and saturates the underlying trackbed, further weakening the structural integrity and accelerating the degradation. A layer of hot mix asphalt within the track substructure was proposed in track renewal for strengthening and waterproofing.

Long-term settlement measurements of several renewed crossings were documented, showing the better performance of the asphalt underlaid crossings (Rose, 2011). However, the asphalt enhancement at the crossings generally does not influence the vertical profile of adjacent granular track approaches, where the approaches eventually settled more than the crossing areas and this differential settlement is potentially a disturbance to the vehicle-track interaction. Therefore, a balanced crossing design is necessary that provides gradual and smooth transitions to both roadway and highway approaches.

Apart from the field measurement, many researchers have established mechanistic models to describe the vehicle-track dynamics when trains pass through the approaches to point assets. Mechanistic models are formulated based on mechanical properties and layout of all the components that make up the track structure and vehicles (de Man, 2002). Track components can be divided based on their principal properties, i.e. the components with mass and inertia properties (e.g. rails and sleepers), those with elastic properties (e.g. railpads) or both (e.g. ballast). Together with the track design, these mechanical properties formulate the relationship between the forces exerting on the track and track responses concerning forces, stresses and displacement, suitable for structural analysis (de Man, 2002). Two paths of modelling solutions are recognised here: analytical modelling and numerical modelling.

2.1. Methods in modelling vehicle-track dynamic interaction

Analytical modelling is suitable to solve problems in continuous support condition with a limited number of connections and loading positions, which facilitates the retrieval of closed-form solutions to track responses (de Man, 2002). Due to a sudden change in the structural solution and/or geotechnical foundations, the track transitions to the point assets are prone to strong amplification of the responses...
and localised degradation. First insight into the induced vibrations caused by the inhomogeneity in track properties (e.g. stiffness) can be obtained by simplified models, where some researchers have proposed analytical or semi-analytical solutions, cf. (Dimitrovová & Varandas, 2009; Färågåu, Mazilu, Metrikine, Lu, & van Dalen, 2020; Färågåu, Metrikine, & van Dalen, 2019; Sadri, Lu, & Steenbergen, 2019; Sadri & Steenbergen, 2018). These models are computationally efficient and able to provide better insight into the track responses. However, they are often mathematically difficult especially when considering the track inhomogeneity and are hard to be validated by field experiments.

Numerical modelling can simulate complex non-homogenous geometry with different levels of sophistication, which is generally the case of track transitions. It has been extensively used to simulate transient track responses with specific concerns that characterise the features of inhomogeneity, e.g. stiffness longitudinal variations, track geometry irregularities, hanging sleepers. A vehicle-track interaction model is necessary for the dynamic analysis and generally proposed including three subsystems to represent the behaviour of the vehicle, the track, and the interaction between the two.

Types of models to represent a vehicle include moving force model, moving mass model, and moving vehicle-system model, as presented in Figure 3. The moving force model is the simplest and has advantages in computational efficiency; however, dynamic behaviour of trains and its impact on interaction with the track structure are not considered (Färågåu et al., 2020; Zhai et al., 2019). The moving mass model accounts for the mass and inertia of the running vehicle but neglects the vibration absorbing effect of the suspension system. The moving vehicle-system model, established based on the theory of multibody simulation (MBS), can represent the mechanical properties of the vehicles and vary in complexity concerning the vehicle degrees of freedom (DOF).

The vehicle is represented by an assembly of rigid bodies connected by flexible and massless elements (Zhai et al., 2019). The bodies typically include a carbody, two bogies, and four wheelsets. Each of them has a maximum of six DOFs and a simplified vehicle model can be achieved by setting physical constraints according to the simulation purpose (Iwnicki, 2006). To better model the train loads and the interaction with the track in critical areas, recent papers mostly adopt this modelling type.

The characterisation of wheel-rail contact is of significance for analysing the track performance at rail transitions, where the vehicle at some points pass over the unlevelled (track geometry irregularities) and suspended (hanging sleepers) track may cause oscillations and influence the interaction with the track (Iwnicki, 2006). Previous studies mostly use Hertzian contact theory to represent the wheel-rail contact. However, vehicle-track interaction is not included when using the moving force model to represent the vehicle system.

Modelling the track structure can be distinguished from the representation of track components. At the superstructure level, the rails are usually modelled by beam elements, generally Euler-Bernoulli or Timoshenko beams. Beam elements can also represent the sleepers, which alternatively are modelled by mass or solid elements. Spring elements are normally used to model railpads and ballast, while the stress in ballast cannot be accurately calculated as shear forces are generally not considered (Wang & Markine, 2019). More advanced ways to model ballast behaviour have been proposed, such as solid elements and discrete element modelling. The track substructure can be normally represented as rigid (for engineering structures beneath the track), a mass-spring-damper system (Winkler-type), or a 2D/3D continuum.

Fully calibrated numerical models with field investigations have been proposed by many researchers to evaluate the transition performance, and the finite element (FE) method has been extensively used. These models differ in levels of complexity in terms of dimensions, dynamic aspects of the problem, representation of vehicle-track interaction, different support conditions, differential settlement, etc., where two aspects are distinguished:

1. the track dynamic responses during train passage (transient analysis),
2. the static change of the track geometry resulting from repeated loading (long-term analysis).

### 2.2. Short-term performance evaluation

Recurrent track geometry problems in transitions to the point assets drive the need for maintenance. Some commonly reported contributors are stiffness longitudinal variations and (unloaded) differential settlement (Gallage, Dareeju, & Dhanasekar, 2013; Hunt, 1997; Wang & Markine, 2019). Stiffness longitudinal variations produce uneven track deflections under loading and affect how stresses are distributed beneath the track, which induces local settlement and further disturbs the wheel-rail interaction. Many works are dedicated to investigating the impact of stiffness variations on track performance.

Andersen and Nielsen (2003) modelled the track as an Euler-Bernoulli beam resting on a Kelvin foundation with varied vertical support stiffness, where the stiffness was modelled by a stochastic homogeneous field. Similarly, Oscarsson (2002) investigated the influence of randomness in railpad stiffness, ballast stiffness and dynamic ballast-subgrade mass on track responses, where field and laboratory tests were combined to support the stochastic track model. Li and Berggren (2010) used track stiffness data obtained from rolling stiffness measurement vehicle to analyse the effect of stiffness variations on the responses.

Various mitigation measures have been tested to reduce the dynamic amplification based on the impact study of stiffness longitudinal variations, where the key idea is to smoothen the variations. Gallego, Sánchez-Cambronerro, Rivas, and Laguna (2016) developed a 3D FE model configured to enable the calculation of various geometries of cross-section and geotechnical features of materials. The
model was then applied to the key points with abrupt stiffness variations (e.g. bridges, tunnels) on a Spanish railway line to calculate the vertical stiffness value of each point and cross-section designs were proposed to control the variations. Sañudo, Cerrada, Alonso, and dell’Olio (2017) considered a ballasted-slab track connection and investigated the optimal spacing of sleepers at the junction. Shahraki, Warnakulasooriya, and Witt (2015) assessed the behaviour of different transition zones when using longer sleepers, auxiliary rails, and improved subgrade. Similarly, Hu, Zhang, Wen, and Wang (2019) investigated various subgrade filling materials in a 3D FE model for a tunnel-culvert transition to explore the economical materials. Further details on the mitigation measures refer to Indraratna, Sajjad, Ngo, Correia, and Kelly (2019).

Another primary contributor, unloaded differential settlement, is mostly attributed to ballast settlement from fouling and/or degradation (particle compaction and abrasion), and settlement of fill and subgrade layers. This often appears in, e.g. bridge approaches and ballasted-slab track transitions, where the ballasted track at the approaches settles more than the adjacent track on bridge abutments or stiffer slab tracks. This issue is particularly severe in soft soil regions (Coelho, 2011). A field monitoring on a culvert transition in the Netherlands revealed that the track settlement consists of two stages: the ballast initially compacted after maintenance, followed by the second stage with the major settlement from embankment and peat layers (Coelho, 2011). By contrast, settlement measurements on a bridge approach in the US showed that the ballast layer is the primary source of differential settlement (Mishra, Boler, Tutumluer, Hou, & Hyslip, 2017). Similar to the stiffness variations, the differential settlement is also a source of disturbance to the train-track interaction. This leads to rapid elevation of wheels and amplification of wheel-rail forces during train passage, which often results in hanging sleepers, loss of contact between the sleeper and ballast layer when the track is unloaded, further accelerating the local degradation.

Some works evaluated the impact of both contributors on the transition performance and concluded that the differential settlement is more critical than stiffness longitudinal variations (Banimahd, Woodward, Kennedy, & Medero, 2012; Lei & Mao, 2004; Milne, Harkness, Le Pen, & Powrie, 2019; Wang & Markine, 2019). The common ways of incorporating the differential settlement in rail track models include, e.g. reducing the local track vertical stiffness, imposing (assumed) uneven settlement profiles, introducing track geometry measurement records, imposing a transition angle at the rail elevation. They are reviewed and discussed below.

Hunt (1997) studied four cases of a rail transition in semi-infinite beams on a Winkler foundation, namely a rail joint, increasing vertical stiffness, different settlement rates, initial uneven settlement profile between a bridge approach and abutment. The synthesis of the cases resulted in a better transition design based on dynamic sleeper forces. Likewise, various uniform settlement scenarios were imposed in FE models to account for the ballast settlement and study the train-track interaction at bridge transitions (Nicks, 2009; Paixão, Fortunato, & Calçada, 2016b; Wang & Markine, 2019). Lei and Noda (2002) considered the uneven track profiles as Gaussian random processes and incorporated the random irregularities in a 2D FE model. Further in Lei and Mao (2004), the uneven profile at the junction was studied by imposing a transition angle $\alpha$, as shown in Figure 4, and the numerical simulation indicated the irregularity angles significantly influence wheel forces, rail accelerations, and carbody accelerations.

While limited research introduces real track geometry measurement in the numerical modelling, Paixão, Fortunato, and Calçada (2016a) integrated a track geometry record and variations in substructure properties to investigate how the uneven profile and stiffness longitudinal variations influence the train-track interaction at a bridge approach. The shape of the uneven profile was imposed to the top of the ballast layer in an FE model. In Banimahd et al. (2012), the local differential settlement was modelled as equivalent stiffness reduction in ballast to investigate the impact of the transition length and train speed on rail displacements, wheel forces and carbody accelerations.

The numerical models suitable for describing the vehicle-track interaction and transition behaviour are becoming more and more complex. They are limited when applied over extended track lengths considering the computational cost. Milne et al. (2019) utilised track geometry measurements and sleeper deflections to parameterise simulations in a 2D FE model, considering the measured variations in track level, vertical stiffness, and hanging sleepers. Four scenarios were processed to investigate their separated/collective roles on vehicle-track interaction, suggesting the variation in track level is the primary source responsible for vehicle dynamics, while the stiffness longitudinal variation is significant for track behaviour. The results imply forecasting the track performance can be achieved by separately
simulating the vehicle and track behaviour. This decoupling enables a computationally-efficient way to predict the track behaviour at an extended length and facilitates relevant maintenance planning.

Despite the extensive use of the FEM, the discrete element method has been proposed in some works to provide insight into the particle-to-particle nature of load transfer within the ballast layer, which is particularly relevant to the transition-related issues (i.e. ballast settlement from fouling and/or degradation). Tutumluer, Qian, Hashash, Ghaboussi, and Davis (2013) proposed a discrete element model that addresses the particulate nature of varying sized and shaped ballast particles and predicts the magnitude of field ballast settlement under repeated loading. As the track transitions always extend up to tens of metres, using discrete element method to simulate the behaviour of ballast particles and transition as a whole is time-consuming. To solve this issue, Mishra, Qian, Huang, and Tutumluer (2014) proposed an integrated approach where loading profiles simulated from a validated analytical track model were used as input for a discrete element model to predict ballast particle accelerations.

As the hanging sleepers are often associated with the differential settlement, modelling the sleeper-ballast interface is important to study transition-related problems, which can be achieved in various ways. Namura and Suzuki (2007) used a settlement law to estimate the ballast settlement value at each sleeper and adopted it as a threshold at the interface to determine the sleeper/ballast gap. Further, Varandas, Hölscher, and Silva (2016) formulated a piecewise equation regarding the on/off contact between the sleeper and ballast layer to account for the hanging sleepers, which was embedded in a 3D FEM for a culvert transition. The similar way has been found in Mååne et al. (2019). Alternatively, the sleeper and ballast layer can be modelled as solid elements so that their interaction is represented by surface-to-surface contact (Banimahd et al., 2012; Coelho, 2011; Wang & Markine, 2019).

Much effort has focussed into the study of transition performance, and bridge approaches, the sections on either side of abutment being much less stiff than the bridge deck, are qualified as an ideal example of track transitions and have been extensively studied. By contrast, the level crossings have not received much attention in the track numerical modelling, which is also confirmed in Le Pen et al. (2014). The studies above limit the scope to simulating transient dynamic responses under train passage at various levels and locations of track components, providing a thorough understanding of certain aspects of physical mechanisms behind the track degradation. However, the transient analysis results are not directly applicable to represent and predict the evolution of track behaviour, which should be extended as discussed in the following long-term analysis.

2.3. Long-term numerical prediction of track degradation

Numerical modelling can be used to predict the long-term track performance, where much attention has been paid to one-dimensional track geometry degradation, i.e. track settlement. The ballast layer is recognised as the primary contributor (Li, Ekh, & Nielsen, 2016; Mishra et al., 2017; Pita, Teixeira, & Robusté, 2004), and previous effort made in settlement prediction mostly consider the permanent deformation in the ballast layer.

The prediction usually requires a vehicle-track interaction model for transient dynamic analysis to integrate with an empirical equation for permanent ballast deformation, followed by an iterative procedure consisting of two modules. As shown in Figure 5, one is the dynamics calculation module with a vehicle-track interaction model to obtain track responses during train passage, e.g. wheel forces (Fröhling, 1998; Vale & Całçada, 2014), sleeper-ballast contact forces (de Miguel, Lau, & Santos, 2018; Nielsen & Li, 2018), track-subgrade contact stresses (Guo & Zhai, 2018), sleeper deflection (Varandas, Hölscher, & Silva, 2014) and ballast stress (Wang & Markine, 2018).

The numerical simulation tools are mostly the finite element method, with exceptions using the finite difference method (Fröhling, 1998) and multibody simulation (de Miguel et al., 2018). The other is the cumulative settlement calculation module considering the repeated loading, where an empirical settlement equation is coupled with the vehicle-track interaction model by the following calculation procedure: (1) the simulated response is used as input to the settlement equation; (2) the transient dynamic analysis is updated in each iteration step to account for the new state of the track response; (3) the accumulated settlement is calculated through repeated procedures until a certain limit value is reached.

Sato’s settlement law (1995) has been coupled to FE models for ballast settlement prediction at bridge approaches (Nielsen & Li, 2018; Wang & Markine, 2018). Considering that the existing settlement laws cannot handle
varied loading magnitude, Varandas et al. (2014) proposed a new settlement model focussing on the magnitude and evolution of ballast forces resulting from the differential settlement. The empirical model was coupled with an FE model with particular inhomogeneous nature of transitions, i.e. hanging sleepers and nonlinear constitutive behaviour of the ballast. While much effort has been made on predicting the ballast settlement, Guo and Zhai (2018) considered subgrade plastic deformation in slab tracks and developed an FE model to account for track weight and local contact loss between track and subgrade. Track-subgrade contact stresses were simulated and imported to an empirical power model to predict the subgrade settlement.

The coupled models to simulate the long-term track behaviour can be seen as a hybrid approach fusing mechanical responses from physics-based models with empirical relations. The physics-based models map the relationships between mechanical properties and responses, contributing to the fundamental understanding of the subject. However, the use of empirical relations has some limitations. These equations were extrapolated site specifically, mostly depending on the number of loading cycles and/or load magnitude, but not on the ballast and subgrade properties (Dahlberg, 2001). They are limited in accounting for the actual track condition and also being generalised to other sites with varied operational and environmental conditions, especially at transitions to the point assets, where the local degradation process is faster than the open track and the empirical relations for settlement prediction no longer seem to apply. Moreover, the applicability of these empirical equations embedded in the rail track models cannot be validated at the current stage, which bounds the reliability of the prediction results for maintenance planning. However, the settlement prediction is valuable in comparing performance of mitigation measures to reduce the maintenance routines.

3. Macro-level maintenance management

The tasks in macro-level management have a serial nature that builds up from condition measurement, degradation modelling to maintenance planning, which reflects an empirically-driven maintenance regime that transforms from data to feature, feature to insight and insight to decisions. The condition measurement involves regular inspection/
sensor-based condition monitoring and data processing to extract features that characterise the track condition. The degradation modelling develops deterioration curves by using the extracted features to analyse patterns of track deterioration over time, possibly with relations of influencing variables, such as tonnage and train speed, to gain better insight into the asset lifecycle and predict their future condition. Further, maintenance regimes are adapted to reflect the asset needs, and optimisation works can be followed to search for the optimal intervention planning that better reflects operating conditions and ensures safety and service continuity.

3.1. Track condition measurement and characterisation

While the mechanistic models in the micro-level investigation mostly focus on one-dimensional track geometry degradation, i.e. the track settlement, the macro-level analysis quantifies the track condition through a set of track geometry parameters, e.g. longitudinal level, gauge, alignment, cross-level and twist, as shown in Figure 6. Initially, tracks are laid with an ideal geometry condition, defined as a nominal value. Each parameter is associated with a nominal value. Local deformation and wear as a result of usage lead to the deviations from the nominal value. These deviations that develop gradually and reach the thresholds will result in track defects, which is an alert for interventions. Figure 7 presents an example of uneven profiles at a bridge transition.

Track irregularities excite the train-track system. Specifically, deviations in longitudinal level and alignment can be classified into short-wavelength and long-wavelength irregularities, which have specific indications for track defect types. Since the longitudinal level defects are primary sources for the excitation and entail greater relevance to the track degradation, most studies consider them the dominant parameter in evaluating the track geometric quality.

The longitudinal irregularities with wavelengths from 0.03 to 2 metres often develop with wheel rolling defects, e.g. rail corrugation and squats (Arvidsson & Karoumi, 2014). These irregularities are related to the resonant vibration of unsprung masses of the vehicles (e.g. wheelsets) and track itself, characterised by excitation frequencies of 20–1000 Hz (Salvador, Naranjo, Insa, & Teixeira, 2016). Defects with longer wavelengths are mostly related to variations in substructure properties, e.g. settlement and hanging sleepers, the typical issues in transition zones. The resonant vibration is associated with semi-suspended masses (e.g. bogie) to suspended masses (e.g. carbody) of vehicles so that the longer-wavelength defects significantly influence the riding comfort.

Methods for measuring the track geometry generally fall into two categories: track recording vehicle (TRV) and onboard vehicle dynamics measurements. TRV is a mature self-propelled vehicle dedicated to measuring, processing, assessing, and storing track geometry parameters. Depending on the measurement techniques, two main principles are distinguished, i.e. chord and inertial measuring systems:

- The chord method measures the track geometry based on a straight-line chord reference, where the mid-chord amplitude is taken as the measured output. Examples can be found in EM120 of Iran (Mehrali, Esmaeili, & Mohammadzadeh, 2020; Movaghar & Mohammadzadeh, 2020).
- The inertial method requires an inertial system as a reference, e.g. carbody, to measure its relative position with the rail in different dimensions. Examples refer to STRIX and IMV100 in Sweden (Germonpré, Nielsen, Degrande, & Lombaert, 2018; Nielsen, Berggren, Hammar, Jansson, & Bolmsvik, 2020), GI-4 in China (Bai, Liu, Sun, Wang, & Xu, 2015; Liu, Xu, & Wang, 2010), and UFM120 in the Netherlands (Westgeest, Dekker, & Fischer, 2012).

In recent years, more research effort has been made on monitoring the track geometry through onboard measurements (de Rosa et al., 2020). This is more cost-effective and its high-frequency measurements offer opportunities for track geometry data analytics. A comprehensive review on onboard sensors for track geometry measurement can be found in Weston, Roberts, Yeo, and Stewart (2015).

The existing studies on onboard measurements can be categorised into two approaches: model-based and signal-based (de Rosa et al., 2020). Model-based approaches map the mathematical relationships between the input and output signals of dynamic systems, cf. O'Brien, Quirke, Bowe, and Cantero (2018), Odashima, Azami, Naganuma, Mori, and Tsunashima (2017) Strano and Terzo (2019), while signal-based approaches use signal processing, statistical analysis, and recently machine learning techniques on the system response signals to draw conclusions on the input data, cf. de Rosa et al. (2020), Salvador et al. (2016) and Wei, Liu, and Jia (2016). The input signals in this case are track geometry parameters and the responses are vehicle dynamics, measurements taken from axlebox, bogie or carbody. Figure 8 presents the main components of a vehicle system.

Once gathering the data from measurements, evaluating and making decisions on each track geometry parameter per unit length is practically difficult as this results in large volumes of data. Often, track quality index (TQI) is utilised to aggregate various track geometry parameters with wavelength variations. The change of the TQI values provides an aggregate-level picture of individual track segments for asset managers to design interventions, where standard deviation and mean over a defined length and power spectral density are among the standardised TQIs in EN 13848-5 (2008).

Various railroad administrations have tailored their own TQIs to fit the local network characteristics, which can be generally categorised into objective TQI and synthetic/artificial TQI (Lasisi & Attoh-Okine, 2018). The former refers to using individual-parameter measurements to formulate an indicator and addresses a specific aspect of the track geometric quality, e.g. track roughness index (the US) (Sadeghi, 2010) and Canadian TQI (Roghani, 2017). The latter develops a mathematical function that describes the track geometric quality by aggregating all the parameters into one equation, e.g. China railway TQI (Bai et al., 2015; Xu, Sun, 2014). These irregularities are related to the resonant vibration of unsprung masses of the vehicles (e.g. wheelsets) and track itself, characterised by excitation frequencies of 20–1000 Hz (Salvador, Naranjo, Insa, & Teixeira, 2016). Defects with longer wavelengths are mostly related to variations in substructure properties, e.g. settlement and hanging sleepers, the typical issues in transition zones. The resonant vibration is associated with semi-suspended masses (e.g. bogie) to suspended masses (e.g. carbody) of vehicles so that the longer-wavelength defects significantly influence the riding comfort.

Methods for measuring the track geometry generally fall into two categories: track recording vehicle (TRV) and onboard vehicle dynamics measurements. TRV is a mature self-propelled vehicle dedicated to measuring, processing, assessing, and storing track geometry parameters. Depending on the measurement techniques, two main principles are distinguished, i.e. chord and inertial measuring systems:
Liu, & Wang, 2011), Q index from ProRail of the Netherlands and Sweden TQI (Attoh-Okine, 2017). However, they lack consensus in approach, resulting in conflicting inference on the track condition (Movaghar & Mohammadzadeh, 2020). Synthetic TQIs are often dimensionless and lack physical meaning (Lasisi & Attoh-Okine, 2018). Some select specific track geometry parameters while dropping others, or assign subjective weights of the parameters to the synthetic indexes; also, aggregating track geometry measurements for an extended length of track may miss exceptions implying safety risk (Lasisi & Attoh-Okine, 2019).

Some researchers have proposed objective TQIs using unsupervised machine learning to overcome the identified shortcomings. Lasisi and Attoh-Okine (2018) proposed principal component analysis (PCA) to combine 31 track geometry features in a low dimensional form without losing much variability in the data. The extracted three principal components were tested better at predicting the defects than the synthetic TQIs. In Lasisi, Merheb, Zarembski, and Attoh-Okine (2019), they extended their work to employ both PCA and T-stochastic neighbour embedding as dimension reduction techniques on the railway geometry data.

Figure 6. Schematic representation of track geometry parameters.

Figure 7. (a) A bridge transition zone with an uneven profile; (b) Longitudinal level measurement at the location, wherein the blue dotted line indicates the nominal value and red dotted lines are the intervention limits.

Figure 8. Main components in a vehicle system and locations of axlebox, bogie and carbody.
Further, they fused the safety concern into the TQI in order to capture track geometry exceptions in the index (Lasisi & Attoh-Okine, 2019).

Another method in developing novel TQIs is based on statistical analysis. Falamarzi, Moridpour, and Nazem (2019) used Pearson correlation to measure the correlation between existing and previous values of each track geometry parameter, where a strong correlation was identified in gauge and twist so that they were incorporated in TQI development. Sadeghi (2010) investigated the railway geometry data by statistical distributions, where a normal distribution pattern was found best fitting frequency curves of the geometry parameters; new TQIs were developed based on the distribution features. The current deterministic TQIs cannot capture the inherent uncertainty when classifying the track condition against the maintenance thresholds. Movaghar and Mohammadzadeh (2020) proposed a stochastic TQI based on a Bayesian framework to incorporate the uncertainty and this strategy was applied to a 900-km long track line in Iran.

### 3.2. Prediction of track geometry degradation

Once converting the raw data to a proper TQI, a mathematical function that describes the track degradation process and predicts the future state can be formulated to inform the maintenance decision-making. Initially, the settlement is mainly considered the controlling degradation factor in the ballasted tracks, and many researchers developed degradation models describing the ballast settlement due to its major role in the overall track settlement. The quantitative modelling of degradation for granular and porous materials is extremely complex and sensitive to specific material properties so that many settlement relations have been tuned to fit the particular data either from in-situ or laboratory tests (Dahlberg, 2001). A detailed review of these models is presented in Dahlberg (2001). These mathematical formulations are empirical/phenomenological models. Compared with the mechanistic models, they are generally easier to handle but lack of physical interpretations. Renewed interests in these models refer to their coupling with the vehicle-track interaction models for numerical settlement prediction, as mentioned in sub-section 2.3.

Significant improvement in track geometry measurement techniques, especially the onboard measurements, enables the access to large volumes of data reflecting the real track condition, where statistical models have been widely proposed and an emerging research stream applies machine learning tools for predictive analytics of track degradation. Higgins and Liu (2017) and Soleimanmeigouni, Ahmadi, et al. (2016) have reviewed statistical modelling of track geometry degradation. The current review built upon these reviews incorporates the recent literature on machine learning applications and provides a taxonomy based on the methods and modelling purposes to facilitate the model comparison, which is presented in Table 1.

Statistical and machine learning (ML) models are distinguished in data-driven models for track geometry degradation. ML as a sub-field of computer science refers to the ability of a system to learn and improve performance from experience, which is widely understood as methods that analyse data, extract patterns and make predictive analysis from often rich and unwieldy data. ML has its foundation on statistics, but their major difference lies in the volume of data involved (Bzdok, Altman, & Krzywinski, 2018).

Machine learning techniques used in track geometry data analytics can be categorised into unsupervised and supervised learning. In unsupervised learning, data are not labelled and no response variables are observed. The motive is to find hidden patterns in input data, which, as described in sub-section 3.1, is applied to extract objective TQIs from the track geometry data. Clustering analysis and dimension reduction are the primary classes of algorithms. Supervised learning involves observable response variables to guide the learning process. It deals with predictive analytics based on the labelled data for both input and response variables, where classification and regression algorithms are primary groups.

#### 3.2.1. Statistical approach

The first sub-category of statistical approaches characterises the track geometry degradation by deterministic models, where regression techniques have been extensively used to describe the relationships between track geometry degradation and explanatory variables, e.g. time, accumulated tonnage, speed, subsoil type. Its application varies from simple linear regression, exponential regression to multivariate regression. Considering the nonlinear degradation process, Liu et al. (2010); Xu et al. (2011) divided the process into tiny time slots and used least squares regression to approximate the degradation over a time slot. A repeated substitution was made in the process by using updated inspection data, and a family of the estimated regression equations forms a prediction model.

Westgeest et al. (2012) incorporated the effect of the subsoil type, sleeper type, tonnage and engineering structures beneath the track on the track geometry degradation in a multivariate regression model. Further, a log-transformed regression model was proposed to map the degradation with explanatory variables, where a survival model characterising the derailment risk and an optimisation model for maintenance planning were coupled (He, Li, Bhattacharjya, Parikh, & Hampapoor, 2015). Similar approaches have been applied in Guler, Jovanovic, and Evren (2011) and Lyngby (2009) with different operating contexts and influencing variables. Some other studies also adopted linear regression to model the track geometry degradation path but extended the models to link covariates such as tamping effect and spatial dependencies between adjacent track sections to make the models more realistic (Andrade & Teixeira, 2013; Soleimanmeigouni, Ahmadi, Khajehei, & Nissen, 2019).

As the current synthetic TQIs provide an aggregate picture of the track segment condition, isolated track geometry defects that exceed the thresholds may not be captured.
Some studies incorporated the defects to predict corrective maintenance needs, which can be supported by logistic regression, survival analysis and classification algorithms. Classification algorithms as a branch of machine learning are presented in the next sub-section. Logistic regression is part of regression analysis but deals with the binary classification problem. When the degradation path exceeding the threshold is deemed as a defect, track condition can be classified as normal and defected states, and logistic regression by its nature is suitable in solving this problem, cf. Andrade and Teixeira (2014), Khajehei, Ahmadi, Soleimanmeigouni, and Nissen (2019), Sharma, Cui, He, Mohammadi, and Li (2018) and Soleimanmeigouni et al. (2019). Survival analysis is widely used to model the uncertainty in system lifetimes. As a common distribution candidate in survival analysis, Weibull distribution has been used to estimate the probability of track defects (Alemazkoor, Ruppert, & Meidani, 2018).

<table>
<thead>
<tr>
<th>Approach</th>
<th>Subcategory</th>
<th>Example models</th>
<th>Modelling purposes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical approach</strong></td>
<td>Deterministic models</td>
<td>Linear regression, exponential regression</td>
<td>To simply model track geometry degradation w.r.t time/tonnage</td>
<td>Andrade and Teixeira (2013); Liu et al. (2010); Soleimanmeigouni et al. (2018); Soleimanmeigouni et al. (2019); Xu et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>Multivariate regression, logistic regression</td>
<td>To establish a relation between degradation and influencing factors</td>
<td>Guler et al. (2011); He et al. (2015); Lyngby (2009); Westgeest et al. (2012)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Logistic regression, survival analysis</td>
<td>To estimate the probability of occurrence of isolated geometry defects</td>
<td>Alemazkoor et al. (2018); Andrade and Teixeira (2014); Khajehei et al. (2019); Sharma et al. (2018); Soleimanmeigouni et al. (2018); Soleimanmeigouni et al. (2019)</td>
<td></td>
</tr>
<tr>
<td><strong>Probabilistic models</strong></td>
<td>(Continuous) Gamma process, Wiener process; (Discrete) Markov chain</td>
<td>- To capture the uncertainty of track geometry degradation over time - To estimate the time period when: ● the degradation path hits the maintenance thresholds (continuous) ● the condition state transferred to the next state (discrete)</td>
<td>Bai et al. (2015); Galván-Núñez (2017); Meier-Hirmer et al. (2009); Mercier et al. (2012)</td>
<td></td>
</tr>
<tr>
<td><strong>Hybrid models</strong></td>
<td>Linear regression coupled with ARMA; Bayesian framework coupled with regression and conditional autoregressive model</td>
<td>To account for spatial correlation of degradation in adjacent track section</td>
<td>Andrade and Teixeira (2013); Soleimanmeigouni et al. (2018)</td>
<td></td>
</tr>
<tr>
<td><strong>Machine learning approach</strong></td>
<td>Classification</td>
<td>Support vector machine, decision tree, ensemble learning (e.g. random forest), linear discriminant analysis, Naïve Bayes</td>
<td>To predict the track (discretized) state for the next inspection/ to predict the occurrence of geometry defects</td>
<td>Bai et al. (2016); Cárdenas-Gallo et al. (2017); de Rosa et al. (2020); Hu and Liu (2016); Lasisi and Attuh-Okine (2018); Sharma et al. (2018)</td>
</tr>
<tr>
<td></td>
<td>Regression</td>
<td>Decision tree regression, random forest regression, support vector regression Artificial neural networks</td>
<td>To predict the continuous values that are representative of the track condition To predict the track condition by considering complex relationships between independent and dependent variables</td>
<td>Guler (2014); Lee et al. (2018)</td>
</tr>
<tr>
<td></td>
<td>Clustering</td>
<td>Hierarchical clustering, k-means clustering</td>
<td>Group geometry data points according to their similarity to evaluate the effect of interventions on geometry condition</td>
<td>Martey et al. (2017)</td>
</tr>
<tr>
<td></td>
<td>Dimension reduction</td>
<td>Principal component analysis, T-stochastic neighbour embedding</td>
<td>- Reduce geometry data from higher-dimensional space to lower dimensions - Produce objective TQIs to characterise the degradation</td>
<td>Lasisi and Attuh-Okine (2018, 2019); Lasisi et al. (2019); Martey et al. (2017); Sharma et al. (2018)</td>
</tr>
</tbody>
</table>
Track degradation is a stochastic process, affected by heterogeneous factors along the track. The deterministic models only consider the nominal degradation behaviour and the resulting maintenance policies may not be robust in the presence of randomness. The second sub-category of statistical approaches uses probabilistic models to describe the degradation dynamics and account for the inherent uncertainty, using theories from stochastic processes, Bayesian inference, etc.

Gamma process is a stochastic process with independent and non-negative gamma-distributed increments (Van Noortwijk, 2009). The feature determines its applicability to characterise monotonic degradation processes, e.g. track geometry irregularities can only grow without interventions. A gamma process model was proposed to describe the evolution of longitudinal level defects, where a cost model was linked for maintenance optimisation (Meier-Hirmer, Riboulet, Sourget, & Roussignol, 2009). Further, the work was extended by a bivariate gamma process to include alignment in the prediction (Mercier, Meier-Hirmer, & Roussignol, 2012).

Wiener process relaxes the monotonicity, allowing variations in the degradation level caused by interventions and measurement errors. It starts at zero and is continuous in time with independent and Gaussian increments. Galván-Núñez (2017) formulated the degradation path of each track geometry parameter as a Wiener process and the model parameters were estimated by Bayesian inference. The failure time within a maintenance cycle was estimated from the degradation sampling paths. Rather than treating the degradation process as continuous, Bai et al. (2015) classified the track condition into four ranks based on Chinese TQI and described the deterioration process as Markov chains, where transition probabilities between states incorporated tonnage and line horizontal layout as explanatory variables.

The stochastic process-based models are directly linked to maintenance decision making. They are useful in predicting track condition within a maintenance cycle. For continuous processes, the track section is considered as defected when the selected TQI is beyond the predefined threshold, calling for interventions. As shown in Figure 9, the first hitting time when the degradation path exceeds the threshold can be estimated and its inherent uncertainty is quantified by a probability distribution, which contributes to the maintenance knowledge regarding the remaining time before interventions. For discrete processes, the track condition is classified into finite states where each is associated with a maintenance decision. However, the limitation of these Markov processes is their basic working principle, Markovian property, where the future state is only based on the current state, independent of the past state. Besides, the model complexity may restrict their applicability when more track geometry data covering more lines or an entire network is generated in the analysis.

Apart from quantifying the temporal variability of degradation process through the stochastic processes, a few researchers accounted for spatial dependencies of degradation in consecutive track segments, as the neighbouring segments having similar structural and operational features tend to exhibit similar degradation patterns (Andrade & Teixeira, 2013; Soleimanmeigouni, Xiao, et al., 2018). This modelling purpose usually requires a hybrid model combining a regression model with techniques specifically addressing the spatial variability in regression parameters. Therefore, this type of modelling is categorised as hybrid models in Table 1.

### 3.2.2. Machine learning approach

Supervised learning algorithms are mostly applied in forecasting track geometric quality, which can be further categorised into classification and regression problems. The former is applicable when output variables are discrete/categorical and the latter deals with continuous variables (Martey, Ahmed, & Attukokye, 2017).

Classification aims at finding to which category a new instance belongs to, based on knowledge obtained from the training of observed instances. Considering the track condition featured by a TQI is either within or beyond a threshold, the problem is a well-posed binary classification problem (de Rosa et al., 2020), where support vector machine (SVM) and decision tree (DT) are commonly adopted to predict the defect occurrence. They can deal with both regression and classification problems. In classification, SVM separates data into different classes by transforming data into high-dimensional space through a kernel function and dividing them with decision boundaries (de Rosa et al., 2020; Sharma et al., 2018). The kernel trick makes the SVM unique classifier that can map non-separable data into high-dimensional space and make it separable (de Rosa et al., 2020). DT compared with SVM has better interpretability and representation. It is a rooted tree that splits a complex decision into several simpler and more interpretable decisions (Martey et al., 2017). The TQIs are predictors and the classes to be mapped are target variables, formulating a top-down approach to construct a DT.

Hu and Liu (2016), based on Federal Railroad Administration (FRA) regulations, initially proposed SVM to predict the change in defect amplitude considering the effect of track class, traffic volume and inspection intervals. Irregularities in profile, cross-level and dip (the maximum value of track longitudinal levelling within a certain length) were involved in the model training. More recently, Cárdenas-Gallo, Sarmiento, Morales, Bolívar, and Akhavan-Tabatabaei (2017) considered the same types of defects and constructed an ensemble methodology for predicting the occurrence of the more-serious defect.

Sharma et al. (2018) also based on the FRA policy applied random forest, SVM and logistic regression on a record of track geometry defects to predict the defect occurrence. Random forest is an ensemble learning method comprising of multiple DTs during training, where each DT is constructed from a random subset of predictors and random variables. The ensemble aims at correcting the overfitting problem often encountered in training the DT models. A total of 30% of recorded defects were used as test data
and random forest appeared to have the best prediction performance.

While most studies analyse the irregularities in longitudinal level measured from TRV, de Rosa et al. (2020) applied DT and SVM to monitor the irregularities in alignment and cross-level from lateral and roll bogie accelerations. The training phase only used data simulated from an MBS and the trained models were tested against the onboard measurement. Instead of binary classification, Bai, Liu, Sun, Wang, and Wang (2016) discretised the track condition into four rank states based on the Chinese TQI and maintenance policy and proposed a tree-augmented naïve Bayes classifier to forecast the track state for the next inspection.

A few studies have treated the forecasting of track geometric quality as regression problems. Martey et al. (2017) studied a mile of track in the US, where a renewal was conducted during the analysis period. They combined unsupervised and supervised learning on TRV data to estimate the effect of geocell installation on the track geometry condition. The inherent complexity in track geometry degradation is caused by heterogeneous factors. Guler (2014) applied artificial neural networks (ANN) model to map the relationships between the degradation and influencing variables using field data. The influencers involve eight mechanical inputs and four environmental factors, which were mostly treated as dummy variables. The study indicated that ANN is particularly useful in learning complex relationships between track condition and multiple interacting factors related to track design, environment and operation. Another application example is found in Lee, Hwang, Choi, and Kim (2018).

3.3. Maintenance intervention planning

Maintenance planning is considered the final step in the macro-level decision making, which couples the prediction from the degradation modelling and determines when and where to perform the maintenance over a planning horizon. Tamping is considered effective in treating track geometry defects and is widely studied. Several optimisation tools for tamping scheduling have been formulated, such as integer linear programming (Dao, Basten, & Hartmann, 2018), mixed-integer linear/nonlinear programming (Famurewa, Xin, Rantatalo, & Kumar, 2015; Gustavsson, 2015; Khajehei et al., 2020; Vale & Ribeiro, 2014; Wen, Li, & Salling, 2016), and heuristic methods (Khajehei et al., 2020; Zhang, Andrews, & Wang, 2013). The parameters of interest include cost, possession time, the total number of tamping operations over a planning horizon, and track condition captured by TQIs. Another important aspect embedded in the optimisation is to model the tamping effect on the track geometric quality, i.e. changes in both the degradation level and degradation rate, as tamping is imperfect maintenance.

However, the local track geometry issues in point assets may not be effectively rectified by tamping. As evidenced in Le Pen et al. (2014), hanging sleepers reoccur soon after the tamping applied at a level crossing approach. Results from investigating the dependency between track geometric quality and longitudinal variations of vertical stiffness in Nielsen et al. (2020); Roghani and Hendry (2017) also confirmed the observation: uneven profiles with high degradation rates often occur on track sections with a combination of a high gradient and low magnitude in substructure stiffness. Further, Yurlov, Zarembski, Attoh-Okine, Palese, and Thompson (2019) linked the occurrence of track geometry defects with subgrade parameters measured by ground-penetrating radar (GPR).

GPR is a continuous non-destructive testing method that measures layer configuration, moisture content and fouling condition to provide a detailed picture of the ballast and substructure conditions. The results revealed a significant relationship between high rates of track geometry degradation and poor track subsurface conditions. This is similar to the case in the level crossings, where, in addition to the stiffness longitudinal variations, the approaches are susceptible to fouling and drainage issues (Shang, van den Boomen, de Man, & Wolfert, 2019). In this case, tamping
may not be a cost-effective long-term solution and upgrading the ballast or subgrade layer should be considered to solve the problem.

Another intervention dealing with the local degradation in level crossings is replacement. Normally the replacement of rail and the corresponding fastening components at crossing areas is associated with the ballast renewal. Rose (2011) and Le Pen et al. (2014) have set up instrumentation to assess their performance before and after the renewal. Few studies have dedicated to optimal maintenance planning on the level crossings. Shang et al. (2019) developed a reliability-based lifecycle costing model for the embedded rails in level crossings. The modelling shows the optimal timing of replacement between the actual rail reliability profile, financial parameters, and maintenance policies. However, the track geometry degradation at the approaches is not included.

4. Perspectives and future direction

The review follows the bottom-up approach aiming to provide a complete picture of which perspectives are related to the maintenance management for the discrete railway assets, especially the level crossings, what types of approaches have been proposed, which techniques are used to develop the models, and what is missing in terms of the efficient management for multiple point assets. The responses to these questions are summarised below, and the research gaps that pave the way to the future research direction are also elaborated.

The model development in the two levels of asset management generally shares the modelling sequence: selecting input and output variables, formulating a functional relationship between the variables and parametrising the model:

- The micro-level investigation mainly adopts the mechanistic approach, where the choice of the variables and form of the model are based on first principles. Detailed knowledge of site-specific conditions and a wide range of track and vehicle parameters are required as input variables. Track responses used to evaluate the dynamic track performance are often axle accelerations, wheel forces, sleeper deflection, etc. The model calibration is normally based on small-scale trackside experimentation.
- The macro-level management process essentially adopts a data-driven approach where all the three tasks are handled by historical data, taken from large-scale network measurements gathered through TRV/sensor-based monitoring. Influencers treated as inputs can be summarised as endogenous and exogenous variables: the former is related to track configuration in constituting components and the latter considers the impact from operational characteristics, maintenance regime and environmental conditions. The outputs are TQIs that capture the track condition over time.

A comparison of the two approaches is summarised in Table 2. Mechanistic models may consider short- and long-term analyses. Short-term performance evaluation is useful in investigating the track responses at various levels and locations of track components during train passage. They can adapt to complex site situations such as variation in track constituting properties, train loading, and velocity. Another strength is that one single model can work out multiple design options through parametric studies, lending itself to testing various track design solutions to optimise track performance and reduce maintenance routines.

The long-term numerical prediction mostly concentrates on one-dimensional evolution of the railway geometry, settlement, embedding an empirical settlement law to support the iterative computation procedure. However, the mechanistic models are applied at the individual asset level. They are limited in accounting for the spatial and temporal variability in degradation, hampering the accurate prediction of the track settlement and generalisation of the results to other sites of interest.

Data-driven models have better predictive capability. The track geometric quality of a 200-m track segment is often aggregated to evaluate the overall condition of a track line or entire network, and the threshold exceedance-based policy is adopted to support tamping operations. Probabilistic models can treat the uncertainty in track degradation; multivariate regression or more advanced ANN models can reveal complex relationships between track performance and exogenous variables, which cannot be coped with by the mechanistic approach. However, the prediction accuracy of the data-driven models depends on the quality and quantity of historical data. These models are blind to physical sources of degradation and unable to account for internal factors related to the structure itself (Steenbergen, 2013).

When reflecting these perspectives on the level crossing assets, the following remarks are considered necessary for future direction. In the micro-level analysis, apart from several field experimentation cases at level crossings, little concern has been engaged to the track dynamic analysis at transitions to level crossings. There are several ways to carefully design transitions onto bridges and over culverts to minimise the dynamic loading. However, there are no recognised transition designs for level crossings, which is also confirmed in Le Pen et al. (2014). The level crossings are common areas where ballasted tracks meet slab tracks (e.g. the use of embedded rail system in the crossing). The optimal design for both the crossing and transition areas is seldom studied and synchronised. Also, as the level crossings represent jointly used areas by rail and road traffic, a balanced crossing design is necessary that provides gradual and smooth transitions to both the roadway and highway approaches. The potential problems caused by the lack of an effective track design at level crossings could be alleviated with an alternative design solution, where more field
experiments and numerical studies dedicated to the level crossings are suggested.

From a macro perspective, it was found that the mainstream in existing railway geometry data analytics focuses on the degradation of open tracks while neglecting the localised degradation features in the point assets, which is reflected in the selection of TQI and maintenance type:

- **TQI:** the intervention planning is mostly supported by TQI-based trend analysis, where the standard deviation of longitudinal level (in wavelength 3–25 m) over a 200-m track segment is a decisive factor. The aggregate TQI however may not capture the localised degradation feature or highlight the higher degradation rates in the point assets as these assets normally extend up to a few metres or tens of metres. Practically, the inspection on these point assets especially the level crossings is often made by regular manual checks (Shang et al., 2019). There is a need to convert the track geometry data to specific track features at point assets to inform local attention.

- **Maintenance type:** the TQI-based trend analysis is generally used for tamping optimisation, mostly coupled with a condition recovery model and replacement concern. Tamping is effective in packing the ballast layer but may not help correct the track geometry defects and hanging sleepers at approaches to the point assets, where the root cause mostly lies in the substructure level. Upgrading the ballast or subgrade layer should be considered as an alternative.

The question which interventions to undertake at the point assets necessitates the proper evaluation of track substructure condition, which requires the synchronised measurement of track layout and stiffness. The stiffness measurement is effective to reveal the potential substructure-related problems. Since filtering track profile data in different wavelengths has specific indications about types of defects, the filtering results can be synchronised with stiffness measurement to refine the defect diagnosis along the track lines, where localised degradation features at point assets are more likely to be captured without manual inspection and interventions specific to the identified issues can be followed. A more advanced way is to synthesise GPR measurement with railway geometry and stiffness measurement as GPR provides additional valuable insight into the substructure condition related to the moisture and fouling condition.

While existing TQI may not capture the defect-proneness of the point assets, hybrid TQI derived from the combined measurement is suggested as it can ease the track characterisation and provide a more precise detection of the poor track condition for proper maintenance treatment. Feature extraction techniques can be applied to define TQI: PCA and T-stochastic neighbour embedding have been tested in mining the railway track geometry data for the open track, and many other methods are underexplored.

As wavelength contents of track geometry defects are inherently related to the specific issues of vehicle-track interaction, full-spectra track geometry filtering is suggested to be incorporated in the TQI development and degradation modelling in order to avoid omission of potential types of defects, where most studies only considered the wavelength range 3–25 m. Besides, the evolution of track geometry irregularities is mainly investigated in the time domain, and application of spectral analysis is not as widely used as the time-domain methods. Since the spectral analysis reveals frequency components in the track geometry defects, it provides a better understanding of the vehicle-track interaction mechanism, showing a great potential to link to the rail track models.

Further, the hybrid TQI that refines the defect diagnosis can be linked to degradation modelling and maintenance decision support and extended from the individual asset level to the system level. Deterministic models are not recommended in the degradation modelling as this type of models only captures the nominal degradation dynamics and the resulting maintenance policies may not be robust enough in the presence of randomness, e.g. the heterogeneity in asset features, degradation levels, and operating

---

**Table 2. Comparison of mechanistic and data-driven approach.**

<table>
<thead>
<tr>
<th></th>
<th>Mechanistic model</th>
<th>Data-driven model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aim</strong></td>
<td>- To provide physical interpretations on structural degradation</td>
<td>- To predict track degradation for maintenance planning &amp; optimization</td>
</tr>
<tr>
<td></td>
<td>- To analyse the impact of mechanical properties and operational variables on track degradation</td>
<td>- To discover complex relationship between heterogenous influencing factors and track degradation</td>
</tr>
<tr>
<td></td>
<td>- To improve the engineering design &amp; maintenance measures at an individual asset level</td>
<td>- To capture the inherent uncertainty of track degradation in the modelling</td>
</tr>
<tr>
<td><strong>Required data/information</strong></td>
<td>- Site specific measurement: accelerometer, deflectometer, geophones, digital image correlation, etc.</td>
<td>- Network measurement: track geometry measurement (TRV or onboard sensors)</td>
</tr>
<tr>
<td></td>
<td>- Track &amp; vehicle design parameters</td>
<td>- Maintenance history</td>
</tr>
<tr>
<td></td>
<td>- Operational characteristics</td>
<td>- Operational characteristics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Environmental conditions</td>
</tr>
<tr>
<td><strong>Approach</strong></td>
<td>- Analytical modelling</td>
<td>- Statistical approach</td>
</tr>
<tr>
<td></td>
<td>- Numerical modelling</td>
<td>- Machine learning approach</td>
</tr>
<tr>
<td><strong>Applicable context</strong></td>
<td>- Mostly transient analysis to investigate dynamic track responses under train loading</td>
<td>Long-term analysis for track geometry defect prediction and maintenance planning</td>
</tr>
<tr>
<td></td>
<td>- Long-term analysis where a numerical model is coupled with an empirical model for settlement prediction</td>
<td></td>
</tr>
</tbody>
</table>

---

---
context, considering the large quantity of point assets on a network. Group maintenance for multiple point assets with similar conditions on the network can be added in this context by using optimisation tools. Economic dependence encompassing the set-up cost (sharing) dependence and operational downtime dependence over multiple assets is embedded as a potential benefit in system-level maintenance decision making.

5. Conclusions

As part of the network infrastructure, the discrete railway assets play a supporting role in the network functioning. Compared with the open track, they are limited in length, but their attributes of large quantity, scattered locations, localised degradation, and potential heterogeneity determine their critical role in network railway maintenance management. The current review synthesises different perspectives related to maintenance management of the point assets, and, as little concern has been engaged to the level crossings, the review adds a new dimension and emphasises on providing a solution to apply maintenance on multiple level crossing assets to ensure asset reliability and service continuity at the network level, which is also applicable for other discrete types of railway assets.

The systems-thinking way has been engaged in the synthesis and two levels of asset management are used to structure the relevant studies. The mechanistic and data-driven approaches shed light on different aspects of the degradation and both contribute to the maintenance knowledge: the former is useful in developing countermeasures to solve local issues such as differential settlement and reduce the maintenance routines; the latter strives for optimal planning of regular interventions, where tamping has been the most studied maintenance operation.

Unlike bridges and culverts, the optimal design for the level crossings and transition areas is seldom studied and synchronised. Besides, for the regular intervention, tamping may not help solve the local defects at the point assets, where ballast upgrading or other measures are alternatives to testing and comparing. This necessitates the synchronised measurement and refined diagnosis of the track issues, especially at the substructure level. Hybrid TQI derived from the combined measurement is suggested. This can be linked with maintenance decision support and further extended to network maintenance planning to ease and optimise the management process. Relevant techniques that can be used in this process are also elaborated.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Chinese Scholarship Council under Grant [201907720116].

ORCID

Yue Shang http://orcid.org/0000-0003-3786-7014
Maria Nogal http://orcid.org/0000-0001-5405-0626
Haoyu Wang http://orcid.org/0000-0003-3131-8217

References


