Canadian Journal of Civil Engineering

Hybrid Rail Track Quality Analysis using Nonlinear Dimension Reduction Technique with Machine Learning

Journal:	<i>Canadian Journal of Civil Engineering</i>
Manuscript ID	cjce-2019-0832.R2
Manuscript Type:	Article
Date Submitted by the Author:	18-Dec-2020
Complete List of Authors:	Lasisi, Ahmed; University of Delaware, Civil & Environmental Engineering; University of Manitoba Faculty of Engineering, Civil Engineering; University of Lagos, Civil & Environmental Engineering Attoh-Okine, Nii; University of Delaware, Civil & Environmental Engineering
Keyword:	Track Quality Index, T-Stochastic Neighbor Embedding, Unsupervised Learning, Railway Engineering, Data Analytics
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

SCHOLARONE™
Manuscripts

Hybrid Rail Track Quality Analysis using Nonlinear Dimension Reduction Technique with Machine Learning

Ahmed Lasisi^{a,1} and Nii Attah-Okine^b

^aPost-Doctoral Fellow, Department of Civil Engineering, University of Manitoba. Mail: E1-310 EITC, 15 Gillson Street, Winnipeg MB R3T 5V6. Email: aolasisi@udel.edu

^bProfessor, Department of Civil and Environmental Engineering, University of Delaware. Mail: 127 The Green, 301 DuPont Hall, Newark, DE 19716. Email: okine@udel.edu

ABSTRACT

Track Geometry parameters from rail track inspection are regulated within unique safety limits for different track classes. This paper focuses on developing an index that combines safety and track quality because of the inefficiency of having corrective maintenance activities between routine maintenance cycles when federal geometry limits are violated. This combination is achievable by summarizing multivariate track geometry parameters, as an improvement to previous linear approaches to address the problem of inefficient track geometry maintenance programs. The use of nonlinear dimension reduction (T-Stochastic Neighbor Embedding-T-SNE) for Hybrid Track Quality Index development, and the influence of time-based parameters on track quality is evaluated in this study. Results show that probability of geometry defects are correlated with principal components but T-SNE had the best prediction on train-test splits despite its poor performance on a blind validation set. The absence of observable correlation between track geometry and acceleration data calls for further investigation.

¹Corresponding Author

INTRODUCTION

Track geometry-related accidents remain a source of concern for railroads and safety professionals. Track-caused accidents constitute 30% to 40% of all railway accidents in the United States over the past decade (FRA 2018), equivalent to \$1.1 billion in financial costs. As railroads continue to invest in the development of Track Quality Indices (TQI), track safety is often treated in isolation despite its obvious interconnection with track quality (Galván-Núñez and Attoh-Okine 2018).

TQIs are designed to measure objective rail track condition and plan maintenance (Lasisi and Attoh-Okine 2018). If overall track quality is assessed through TQIs without reflecting the possibility of a threshold violation for a wide gage or degraded surface profile, then the purpose of TQI needs re-examination. That is, if geometry maintenance regimes are planned with TQIs, why are TQIs not sensitive to safety exceptions? Hence, what is the basis for planning maintenance with TQIs? Therefore, this paper aims to develop an overall TQI that characterizes track degradation by combining track geometry parameters and vibration data with critical thresholds that have safety implications. As a result, the use of such TQIs for maintenance will be justified.

Due to foot-by-foot measurements and per-unit length hypersensitivities of raw track geometry data (Mohammadi et al. 2019), it is practically convenient to aggregate track data into discrete segment lengths, with measures like standard deviation, power spectral density and variance. This is one dimension of uncertainty in track geometry data analysis. Other dimensions arise when different parameter TQIs are expressed under unique wavelengths, e.g. TQI in 31ft, 62ft, 79ft and 124ft for Surface Left and Right, Alignment Left and Right, Gage, Cross level, Warp and Cant. These combinations make the total number of track geometry parameters grow rapidly. Of all these combinations, researchers have sought which TQI representation reflects overall track degradation. Previous studies suggest aggregated TQIs using weighted sums (Alfredo

and Corbin 1986; El-Sibaie and Zhang 2014). Lasisi and Attoh-Okine (2018) posited that the weight assignment for each track geometry parameter/TQI is subjective. The study discovered that combined TQIs are likely to misrepresent the intended measurement (e.g., a TQI that measures horizontal degradation instead of overall track degradation) if inadequately calibrated. Other researchers described the purpose of each track geometry parameter and recommended their respective use cases. For example, Soleimanmeigouni et al. (2018) suggested that tamping maintenance cycles are best predicted using Surface TQI.

Figure 1a describes the state of track geometry maintenance practice. The appropriate TQI is selected as a surrogate of deterioration while cumulative loading/Million-Gross-Ton (MGT) represents track usage over time. A track continues to deteriorate as the TQI increases based on increasing MGT. Routine maintenance is scheduled according to a pre-defined TQI threshold. Safety thresholds set by federal regulators (e.g. FRA) are based on individual track parameters, not TQI. Hence, safety thresholds can be exceeded between two routine maintenance cycles. This necessitates a slow-down order or closure until the exception is fixed. This can negatively impact overall network velocity (Li et al. 2014) and increase maintenance costs. Figure 1b proposes a hybrid index that combines all geometry parameters through the application of data science in a way that track quality and safety are considered. It eliminates the need for spot or corrective maintenance in-between routine maintenance cycles. Another advantage of this approach is that fewer maintenance cycles lower the rate of deterioration and maximizes track uptime (Nielsen et al. 2018). This will consequently minimize track maintenance cost and disruptions. To achieve this, a multi-disciplinary data science and machine learning approach is required. The following subsection summarizes machine learning applications in rail track engineering relevant to this study.

Data Science and Machine Learning in Rail Track Engineering

Despite the plethora of work in railway data science applications, Table D1 summarizes the most crucial literatures to this study. Emphasis is made mostly on track and geometry maintenance. This does not imply the paucity of literature on rail defects because such works are limited in Table D1. Ghofrani et al. (2018) details recent big data railway applications while Attoh-Okine (2017) is a comprehensive resource for machine learning techniques and use cases.

While track geometry maintenance works are based on track quality indices, the veracity of those indices require advanced inquiry (Alfredo and Corbin 1986; Lasisi and Attoh-Okine 2018). This paper aims to bridge that gap by taking advantage of recent advances in nonlinear dimension reduction to objectively streamline multivariate track geometry data for effective maintenance decision making. A forerunner of this study (Lasisi and Attoh-Okine 2018) summarized the geometric quality of rail track in three dimensions. Figure 2 presents a threshold case for Gage using the third principal component (PC3). The values of PC3 reflect safety and track quality. On the y-axis, TQI is standardized with a mean of 0 and standard deviation of 1. The true values of Gage for the inspection date (08-15) heralds the defect picked up on (11-15). Therefore, a threshold value set at 1.65 would preempt a maintenance call on this property. Values of other TQIs like the Polish J-coefficient and track geometry index (TGI) were also low for that inspection date. Similar analysis is possible for surface and alignment.

MAINTENANCE AND TRACK QUALITY INDEX (TQI)

Maintenance is necessary to keep railway track functional (Esveld 2016). Maintenance activities involve the use of specialized equipment and machinery to repair and renew track substructure and superstructure. Examples include surface corrections (tamping, stone-blowing and stabilizer), grinding, ballast renewal, cleaning and undercutting as well as drainage maintenance (Dingqing et al. 2016).

Maintenance Approaches

Maintenance can be corrective or preventive. Recently, data-driven maintenance practices are popular because of the ubiquitous nature of data collection systems and tools (Soleimanmeigouni et al. 2018). Corrective Maintenance is carried out when track geometry parameter exceeds safety thresholds (FRA 2002). Track geometry deviations are fixable by tamping. Tamping significantly reduces the rate of track deterioration by leveling track geometry irregularities (Martey et al. 2017). Derailments have occurred subsequent to tamping operations if track is not properly settled (Dingqing et al. 2016). Corrective Maintenance is not the most efficient approach to maintenance in terms of cost, resources and time (Sharma et al. 2018).

Preventive maintenance on the other hand, is the first step towards data-driven track maintenance (Su and Schutter 2018). It estimates degradation rates to provide directives for intervention without violating critical thresholds. It should also not prescribe early maintenance that squanders limited resources (Su and Schutter 2018). A measurable outcome of this approach is cost reduction and safety improvement (e.g., lower annual derailments). While cost reduction can be monitored in the short term, it would take longer time for safety improvements to manifest for a large railway network (Zhang et al. 2013).

Data-Driven Maintenance

With the growing need for data-driven decision-making, railroads are continually pressured to make the best of collected data. A data-driven maintenance approach transcends forecasting rail track defects with robust analytical tools. It is a holistic process that involves inspection planning, data collection, analysis, alternatives evaluation, policy planning, benefit-cost analysis amongst others. Figure 3 shows a process that begins with collection of inspection data and ends with updating existing policy. Inspection is mandated by FRA (2002) and sometimes scheduled internally by large railroad organizations. With minimal processing, exception or violations are identified from inspection data for spot defect rectification. Examples include, spot geometry

corrections, tie replacement or fastener adjustments. One step further is to process inspection data and conduct prognostic analysis on multi-feature track information. Common data analytical tools employed include timeseries analysis, supervised and unsupervised learning algorithms, deterioration modeling and other methods later discussed in this paper.

From Figure 3, geometry inspection data is used to compute TQIs and rail defect growth rates that are fed into a Markov Decision process for the purpose of defect forecasting. At this stage, macro-level decision analysis can be considered for a large infrastructure network before deploying optimization algorithms for decision processes. The result is an ideal but impractical policy. This Ideal Policy is updated into a Real Optimal Policy that is fed into the existing policy. A continuous evaluation of maintenance policy is conducted in terms of cost, benefit and other metrics with adjustments continually fed into the decision process. The whole process continues until the cost reduction objective approaches convergence or a new objective function is defined. This approach considers only track geometry and rail defects. Therefore, this study focuses on TQI development through nonlinear dimension reduction as a crucial component of the data-driven maintenance approach illustrated in Figure 3.

Individual TQIs

Individual TQIs, also known as objective TQIs aggregate each track geometry parameter over a fixed length. Objective TQIs are described in Sadeghi (2010) and Lasisi & Attah-Okine (2018), while a visual illustration of crucial TQI-based parameters is provided in Figure 4. Other track quality indices related to track substructure and their effects on track geometry integrity have been studied by Dingqing et al. (2016) and Sadeghi et al. (2019).

Combined TQIs

Combined TQIs are predicated on the assumption that a single parameter can effectively represent overall track quality. TQIs of this form have been developed by computing a weighted

sum of each track parameter. Weights subjectively assigned to unique parameters can lead to inconclusive interpretations of overall computed values (Lasisi and Atttoh-Okine 2018). Unlike weighted sums, two types of combined TQIs described below take a different approach. Others have been discussed in previous studies (Sadeghi 2010; Galván-Núñez and Atttoh-Okine 2018; Lasisi et al. 2019).

Australian Rail Track Corporation (ARTC) TQI

Also known as Five Parameter Track Defectiveness, it can be expressed as:

$$W_5 = 1 - [(1 - w_{profile})(1 - w_{align})(1 - w_{gage})(1 - w_{cant})(1 - w_{twist})] \quad (1)$$

where w is the calculated defectiveness for each geometry parameter.

While ARTC TQI shows no explicit weight assignments, it implicitly assumes that all parameters have equal importance. Secondly, this TQI is not mathematically robust because it is susceptible to vanishing products because it assumes an "AND" fault tree relationship. A track section could be defective either through one of the parameters or via several combinations. In other words, any parameter with a defectiveness of 1, simply reduces the whole product to 0.

Swedish Q Index

$$Q = 150 - 100 \left[\frac{\sigma_H}{\sigma_{H_{lim}}} + 2 \frac{\sigma_s}{\sigma_{s_{lim}}} \right] \quad (2)$$

where σ_H is the average standard deviations of right and left level, σ_s is the average standard deviations of alignment and cant while $\sigma_{s_{lim}}$ and $\sigma_{H_{lim}}$ are the threshold TQI values (Muinde 2018).

The Swedish Q Index expresses a 100% superiority of longitudinal level over profile. Clearly, other parameters have been excluded/discounted. This makes it short of a robust combined TQI. As a result, no rigorous analysis will be conducted on ARTC TQI and Swedish Q Index in this study.

Alternatively, a data-driven hybrid approach that allows for the objective combination of all track parameters based on individual variability (and local data structure) is proposed in this study.

This allows for the inculation of varying wavelengths of parameters like surface, alignment, and warp. Other examples like the Austrian TQI, Indian GQI, Polish J index, etc. have been the object of reputable studies (Hamid and Gross 1981; Sadeghi 2010). This study focuses on unraveling latent information resident in the n -dimensional space of multivariate track geometry data because most of the afore stated studies and indices are based on mechanistic measurements with simple statistical summaries. Hence, the proposed approach employs linear and nonlinear dimension reduction techniques to project track geometry data onto a hyperplane where most of the variance exist.

STUDY FRAMEWORK

The first stage of this work featured unsupervised machine learning (principal component and T-SNE) for aggregating multi-dimensional track geometry data. Track geometry inspection data was obtained from two different sources. While both sources are from Class I U.S. railroads, one is a passenger railroad (DAT1) and the other is freight (DAT2). Lasisi & Attoh-Okine (2018) studied TQI development for a passenger railroad using linear dimension reduction. This study focuses on freight railroad track geometry data using both linear and nonlinear dimension reduction. The data obtained from the freight railroad contains two databases: time-based and distance-based track geometry data. These databases cover track geometry inspection measurements for more than 50 miles of track. The time-based data contains information on 3D vehicle accelerations and rotation. These variables are important for ride quality especially for passenger operations.

Exploratory data analysis was conducted on DAT1 and DAT2 data sets to assess latent trends in independent variables before unsupervised learning commenced.

The application of a popular linear dimension reduction technique was implemented on DAT1 to summarize track parameters into three (3) components. The verification of these components for maintenance applications were subsequently conducted. Standard supervised machine learning methods were selected for this exercise. DAT2 was analyzed almost identically because of the

nature of the data structure (2 databases). Firstly, objective track quality indices were developed from distance-based data while an exploratory data analysis (EDA) was carried out on time-based measurements. A subsequent analysis eliminated redundant acceleration parameters and gyroscope readings based on the outcome of EDA and TQI development. Embedding (nonlinear dimension reduction, NDR) operations were then implemented to reduce multivariate distance-based measurements into two or three critical components. The last phase of DAT2 analysis extracted probabilities of geometry defects based on observed geometry data. The emergence of geometry defects is then observed from low to high probabilities, with the corresponding principal and embedding components. Based on the outcome of this correlation, the probability of defect or artificial component (Hybrid Index) can sufficiently track a specific or overall track quality as shown later. This methodology replaces the subjective biases of weight assignment hitherto employed in rail track TQI development (Alfredo and Corbin 1986) with a data-driven approach.

Unsupervised Machine Learning

Unsupervised machine learning deals with a class of statistical learning that unravels hidden critical relationships in data. With no target variables in mind, it includes such methods as clustering, dimension reduction and pattern recognition. Supervised learning on the other hand aims to learn a function from a set of independent variables in order to predict a target variable (Lasisi and Atttoh-Okine 2019b). For the purpose of this study, a Scikit-learn implementation of multi-layer perceptron (MLP) has been selected as the supervised learning technique (Pedregosa, Varoquaux, and Gramfort 2011) to predict the presence of track geometry defect using a selection of predictors from track geometry, acceleration data and artificial indices.

Linear Dimension Reduction

The artificial indices employed are created from linear or nonlinear reduced dimensions. Dimension reduction and feature extraction are interchangeable terminologies. It is the idea that data can be projected onto hyperplanes (or principal axes) without losing relevant information in the dataset.

Principal component analysis (PCA), principal factor analysis, maximum likelihood factor analysis, projection pursuit, nonlinear PCA, T-SNE are all different dimension reduction techniques, although the last two are nonlinear.

Principal Component Analysis

For the purpose of this study, the description of PCA expressed in track geometry parlance found in Lasisi & Atttoh-Okine (2018) suffices. Before application, over thirty (30) parameters were obtained from geometry-based data and sixty-eight (68) from acceleration/time-based data. These were successfully reduced to three (3) predictors through the application of principal component analysis. Performance evaluation was done using true positive rates, false positive rates, and overall mis-classification rate.

Nonlinear Dimension Reduction

Linear dimension reduction like PCA typically do not keep similar data points close together, especially if the data lies near a low-dimensional space. NDR methods were developed to address this gap and preserve local structure of data. For example, points with similar defects should share neighboring positions different from sections without defect.

T-Stochastic Neighbor Embedding (T-SNE)

T-SNE is one of the popular NDR methods that converts high-dimensional data to a pair of matrixes pairwise similarities used for visualizing inherent similarity in data (Van Der Maaten et al. 2014). It has found application in several fields of machine learning visualizations and image recognition. T-SNE will be described in the context of rail track as follows: Given a set of N track geometry and high frequency acceleration data points, $x_1, \dots, x_N$, T-SNE finds the probability p_{ij} based on the similarity between two inspection observations x_i and x_j . This probability p_{ij} also known as conditional probability $p_{j|i}$ is given as (Van Der Maaten and Hinton 2008):

$$p_{j|i} = \frac{e^{-\|x_i - x_j\|^2 / 2\sigma_i^2}}{\sum_{k \neq i} e^{-\|x_i - x_k\|^2 / 2\sigma_i^2}} \quad (3)$$

p_{ij} describes the similarity of observation point, x_j to another observation x_i . Assuming Gaussian conditions with a data point centered at x_i , x_j would be selected as a neighbor to x_i in proportion to the probability density centered at the latter.

$$p_{ij} = \frac{p_{ij} + p_{ji}}{2N} \quad (4)$$

If $i = j$, then $p_{ii} = 0$. Under Gaussian conditions, the value of σ_i varies with the density of the data.

Hence, smaller values of σ_i are employed in denser parts of the data. Because of the Euclidean distance, $\|x_i - x_j\|$ becomes difficult to differentiate at high dimensions, therefore, p_{ij} converges to a constant, as such, not dissimilar. To solve this, T-SNE, learns a d -dimensional array $y_1, \dots, y_N$ that expresses the similarities p_{ij} in the best possible manner. This similarity is measured as q_{ij} between two mapped points y_i and y_j using an approach similar to q_{ij}

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_k - y_i\|^2)^{-1}} \quad (5)$$

With $q_{ii} = 0$, a heavy tail Student-t distribution (Brereton 2015) is used to measure the similarities between points in the lower dimensional space so as to allow dissimilar objects to be maximally spaced apart. These point locations are determined by minimizing the Kullback-Leibler divergence (Kullback 1978) of the distribution of Q from P :

$$KL(P | Q) = \sum_{k \neq i} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (6)$$

Consequently, a map that expresses the similarities or differences between high dimensional inputs is produced through this optimization. This study employs a Scikit-learn Python implementation for T-SNE as well (Pedregosa, Varoquaux, and Gramfort 2011).

CASE STUDY

Data Description

Data was obtained from a US Class I railroad. About 136 miles of track was inspected with three different runs. This analysis considers only 28 miles of track that was characterized with geometry defects. Several red and yellow defects were recorded and further discussed in the analysis. Pseudo-defect thresholds were created. Since in principle, thresholds are pre-determined. The database comes with two different sets of data: time-based (TBD) and distance-based (DBD). For every, distance-based measurement (every foot), there are about 5 to 20 time-based observations. Figure D3 conveys a visual description of about 722,000 distance-based samples corresponding to over 5,000,000 time-based observations. The time-based observations comprise of accelerations and gyroscope measurements in three dimensions (3D). These measure stability and rotation respectively. Both databases are synchronized with a common index field joined on the DBD (every foot). While both databases collectively have approximately hundred (100) parameters initially, only thirty were identified relevant to track geometry and accelerometer readings, and therefore selected for model training and dimension reduction. Different components have also been identified in the data set including but not limited to switches, turnout, grade crossings, frogs, etc. Selected track geometry parameters (DBD) include super elevation, gage, twist, alignment, surface, cross level, etc. of parameter wavelengths ranging from 31ft, 62ft, 79ft to 124ft. Accelerations and rotations in 3D were among the parameters selected from (TBD).

Distributions

The nature of geometry parameters is further confirmed to be normally distributed or near-normal as previously examined (Lasisi and Atttoh-Okine 2018). Gage and Cant exhibit an almost normal distribution with some skewness to the left while Surface with the different wavelengths is the archetypal Gaussian form (Figure D3). This is necessary to establish a strong basis for T-SNE, a dimension reduction technique that assumes initial Gaussian normality conditions.

Box Plots

Examining track data with boxplots elaborates on the unique variations of each parameter. From Figure D4, it is observable that alignment parameters are all centered at zero because longitudinal deviations are limited. The figure shows that higher wavelengths do not imply higher deviations. This challenges how FRA limits tolerate higher deviations for higher TQI wavelengths (FRA 2002). It again begs the question of TQI effectiveness mentioned previously. Furthermore, time-based parameters exhibit a central tendency around zero, except for vertical acceleration which seems to be moved by a constant phase-shift. Next, feature-to-feature relationships among parameters are considered.

Correlograms and Feature-to-Feature Relationship

Since track geometry parameters are arguably normally distributed, it is safe to examine the inter-correlations between parameters using conventional Pearson's correlation coefficient based on joint normality assumptions (Bishara and Hittner 2017). Figure 6a illustrates the correlation between long and short wavelengths variations of the same parameter (e.g., Surface), and the fact the low and high rail (L and R) parameters of the same wavelength are mostly uncorrelated. This phenomenon questions the basis of averaging the measurements of the left and right rails as seen in developed TQIs (Muinde 2018). The figure also shows a mild correlation between Gage and Alignment, but this is not enough to establish a clear interaction between them. Rather than focus on geometry only, time-based parameter feature-to-feature relationships are important to address the question of track geometry impact on ride quality. Figure 6b shows an obvious lack of correlation between time-based parameters and geometry. Data integrity issues may be an explicable reason for this, other reasons are provided in the section on Discussion and Validation while discussing proposed machine learning model results with different predictors from Table D2.

Principal Components and T-SNE Plots

It is concluded that three (3) components were sufficient to summarize multivariate track geometry parameter (Lasisi and Attoh-Okine 2018). The use of percentage variance as a measure of principal components' ability to summarize data is a cross-disciplinary approach that has been adopted in Engineering, Economics, Health Sciences and other fields outside the domain of statisticians and mathematicians (Bartholomew 2010; Index and Control 2019; Krishnan 2010; Redfern and McLean 2014). This is also clear from Figure D1 where three components already explained over 96% of variance in the data considered in this study. To consider thresholds for artificial components of track data, a low dimensional representation or 2D mapping of the first three Principal and T-SNE components were plotted for a chosen section of the track with several alignment defects and surface. Figure D5a shows the principal components plots and possible threshold information. The surface parameter can be prevented from reaching the defect threshold for all wavelengths if the first principal component is observed to not exceed 10. Similar analogies can be made for other geometry defect types excluding T-SNE components where a linear threshold is impossible because all defect measurements are scattered within non-defective observations. Hence, there is no separating hyper-plane at lower dimensions. To address this, higher dimensions would be considered, or a nonlinear threshold would be created. Both are clearly beyond the scope of this study. Most defects are "covered" by the no-defect observations in T-SNE compared to PC plots (Figure D5). To fully maximize the potential of T-SNE as defect predictors, the best algorithms would have to learn the nonlinear structure of geometry data to create separating hyper-planes for robust multi-class defect prediction.

Defect Predictions and Thresholds

In the United States, safety thresholds are set according to the Federal Railroad Administration (FRA

2002) although a few individual railroads provide maintenance thresholds (or yellow flags) according to internal standards. A red flag is a total violation of at least one FRA threshold. In principle, thresholds are predetermined and not learned from the data. This analysis uses FRA Class Track 4 with over 3,000 defects identified, with Gage being the most represented (see Table D3). Only one defect involving two parameters co-occurred (Gage and alignment).

Model Training and Validation Approach

Based on the foregoing, a machine learning model was trained with tuned hyperparameters to predict defects with the following classes of predictors:

- Distance-based Geometry Parameters
- Time-based Parameters
- Time-Based and Distance-Based
- 3 Principal Components
- 3D T-SNE Components

Training is of the form:

Draft

$$\vec{y} = f(\mathbf{X}) \tag{7}$$

where:

$$\vec{y} = \begin{pmatrix} NO\ DEFECT\ (NONE) \\ GAGE\ DEFECT \\ ALIGNMENT\ DEFECT \\ \vdots \\ \vdots \\ SURFACE\ DEFECT \end{pmatrix} \tag{8}$$

1

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{np} \end{pmatrix} \tag{9}$$

Warp and defect interactions (e.g., Gage and Alignment) have been excluded because of limited defect samples. A multi-layer perceptron (MLP) was trained on different predictors with the hyperparameters in Table D2. Amongst several models initially selected for training, the MLP was the model of choice (Pedregosa, Varoquaux, Gramfort, et al. 2011). A discussion on model selection has been excluded since a similar analysis was done in Lasisi and Atttoh-Okine (2018). However, the findings of this model results are unique. Based on the results from Table D2 it is observed that time-based data (TBD) poorly performed on all defect types. This does not only challenge the suitability of high frequency track data for ride quality assessment but also indicates that it is not a good predictor for track geometry defect. Also, 3D components of T-SNE proves to be consistent as an excellent predictor across all geometry defect types to the point that it outperforms original track geometry parameters on a significant number of occasions. It emphasizes the point that T-SNE preserves the local structure in high-dimensional data through low-dimensional mapping (Van Der Maaten 2009). Interestingly, Figure D5b shows a T-SNE component plot that poorly separates out defect types for a chosen segment of track while MLP predictions show that 3D T-SNE components perform better than other predictors considered in this study. One possible explanation to this anomaly is that the nonlinear structure in the T-SNE representation is effectively learned by a nonlinear algorithm like MLP—an indication that linear learners are appropriate for this problem.

Secondly, principal components performed well, next to 3D T-SNE and DBD. It buttresses the fact that 3 PCs are sufficient to characterize track geometry data. Another property that is further explored in this study is the potential to exploit the linear property of principal components for threshold TQI setting. This point is elaborated later in the discussion.

Furthermore, Table D2 consistently shows the inability of TBD to predict track geometry defects for any class of defect considered in this study. While it may be presumptuous to conclude that

high frequency acceleration data is independent of track geometry data, the disparity in data collection systems for both class of variables may be the underlying cause of the seeming lack of correlation.

Lastly, a blind prediction was conducted on a healthy track segment of about 150,000 data points (or 27 miles) without defects to test the model for over-fitting and false-predictions due to each predictor. Table 1 presents an interesting finding.

Table 1 demonstrates the reliability of 3 Principal components with only 7 alignment false positive data points on a test track of 143,433 observations. T-SNE on the other hand is very likely to predictive false alignment defects despite good prediction performance for a seen or observed track. Table 1 further emphasizes undesirable results due to high frequency data for direct track geometry assessment. Direct track geometry predictor yielded only a handful of defects due to Alignment and Gage. 61 false positives are insignificant when compared to about 150,000 data points.

Defect Probability Relationship with Principal Components

To simplify the idea of principal components and safety threshold, the probabilities of defects for observations with PC1 greater than 10 (see Figure D5) were cross-examined. These probabilities were created using a SoftMax function as the final input to the MLP predictor. A simple framework for this is shown in Figure 7. The SoftMax function is given as:

$$\sigma(\mathbf{x})_j = \frac{e^{x_j}}{\sum_i e^{x_i}} \quad (10)$$

In simple terms, the SoftMax layer is a convenient function that outputs a probability distribution from the MLP output before assigning defect to the parameter with the highest probability. The SoftMax layer is guaranteed to output a probability distribution because even if the outputs of the computational unit is negative, the exponential power of such negative number is positive. Also, it sums to 1 because the denominator sums over all possible values of the numerator (Charniak 2018). To track the relationship between probabilities and artificial track components (PCs), the defects- especially surface-examined from Figure D5 were considered. Figure 8 shows a direct

relationship between probability of Surface defect and 1st principal component. Bearing in mind that these are the only surface defects within the selected section, the probability of defect is not required to be high for the track section to violate pre-determined thresholds. Hence, setting a probability threshold for surface defect could be anywhere between 0 and 1. Although, the subject of optimizing the probability threshold for effective maintenance is a focus of an ongoing study, the developed relationship between PC1 and Surface defects shown in this study can be modeled and characterized by a nonlinear expression that serves the purpose of using PC1 as a track quality index to develop a prediction model. Because such prediction modelling exercises have been studied in previous works (Galván-Núñez & Attah-Okine 2018; Lasisi et al. 2019; Sadeghi 2010), it is considered less expedient in this analysis.

Validation and Discussion

In this section, the validity of the results from the foregoing analysis and implications for rail track maintenance is discussed. The validation process employed in this study analyzes the parameter importance and the predictive power of defect indicators (DBD, TBD, Combined, 3 PCs and 3D T-SNE) through a machine learning process applied on a North American FRA Track Class 4 railroad property. The validation process shows that the False Positive Rates (FPR) in Table D2 are generally low for most predictors except TBD with a correspondingly poor TPR and accuracy. The lack of predictability from the accelerometer readings (TBD) is due to many reasons. One reason is the lack of proper signal processing and collection methods from single and multiple integration errors to integration of random walk, offset drift and inappropriate high-pass filters (Weston et al. 2015). Despite this shortcoming, TPR and FPR values in the combined predictors (Table D2) are lower than those in DBD as a way of verifying that TBD has a lower predictive effect on track geometry irregularity. Also, a blind prediction described in Model Training and Validation Approach was conducted to verify similar conclusions.

Analysis shows that track geometry parameters can be expressed as reduced dimensions, and also as probabilities of defects for each parameter. While the expression of track geometry parameters in terms of defect probabilities could provide a direct intuition for track managers or superintendents, the range of such probabilities require meticulous calibration. For example, Figure D2 shows a range of probability for observations pre-determined to violate alignment safety based on preset thresholds. Firstly, the probability is discrete for about 30 alignment defects even though the figure attempts continuity. Secondly, a track section with alignment defect probability of 0.7 is as equally likely as that with 0.2. Therefore, it is required to study and calibrate the probability with a comprehensive defect dataset before generalizing conclusions and setting probability safety thresholds.

Lastly, the idea of linear artificial track components will require time for adoption. Nonlinear dimension reduction like T-SNE seems to overfit learned data to the point that new false defects were predicted for a blind test on a healthy track segment. However, principal components performed greatly on training and test sets with minimal false positives on the same blind test (0.0049%). Apart from principal component analysis being a linear dimension reduction, it is a classical mathematical method that is easy to conceive and visualize. Its algebraic and geometric dimensions also make it easier to conceptualize. Hence, the suitability of expressing multivariate track geometry parameters with 3 artificial principal components is reasonable and pragmatic.

CONCLUDING REMARKS AND FUTURE WORK

This paper formally described the work on dimension reduction as an alternative to weighted-sum combined track quality indices. This study showed the subjectivity of TQI formulations behind artificial track quality indices while using ARTC TQI and Swedish Q index as examples. An objective data-driven approach was proposed in this study to mitigate subjectivity in TQI development. A previous study hinted at the possibility of using principal components as an

effective linear dimension reduction for multivariate track geometry parameter. This study improved on that by employing a contemporary feature extraction technique: T-SNE. Despite its mathematical robustness, T-SNE has the tendency to overfit the data on seen training and test set but produced tremendous false positives on a blind test using a healthy track section. From a practical railway perspective, treating a couple of critical defects on track is quite pragmatic than creating a technique that identifies about 50% of healthy track with defects. Hence, T-SNE may be well suited for studying known track segments rather than unknown sections.

Secondly, the correlation between time-based observations and track geometry defects was investigated in this study. Through multi-facet exploratory data analysis, this study concludes that there is no apparent relationship between acceleration data and track geometry defects. Through machine learning, the performance of TBD was found to be poor towards the prediction of observed geometry defects-another evidence that TBD could be doubly-integrated to estimate track displacements rather than predict track geometry defects. Although some studies have shown some underlying correlations between the two (Sadeghi et al. 2017, 2020), the accelerometer data collection and processing in those studies are not identical with those used in this paper. Furthermore, accelerometers in such studies were mounted on wagon floors and other vehicle parts which could have dampened the response from wheelsets compared to accelerometer readings collected directly from instrumented wheelsets or wheel frames. Future studies will investigate the effect of damping and source of accelerometer readings to establish an interrelationship between track geometry readings and three-dimensional vehicle acceleration. The proposed study will help railways and track engineers improve their understanding of track geometry data and monitoring from a data-driven perspective. This work is a viable step in assessing the viability of dimension reduction in track geometry data analysis. Future work will also focus on characterizing the probability of each defects with the intention to define objective safety thresholds that reflect true

overall track quality. As an extension to this, a risk vulnerability method will also be inculcated by considering the severity of a track accident at a corresponding probability of defect. In the absence of such data, simulation becomes very important. There is no doubt that the infrastructure systems approach and data science nexus can offer a lot to railroad engineering. This study is a crucial step in that direction.

DATA AVAILABILITY

Some or all data, models, or code generated or used during the study are available from the corresponding author by request (list items).

- Code
- Masked Data

ACKNOWLEDGMENT

The authors would like to acknowledge the University of Nevada, Las Vegas (UNLV/USDOT) University Transportation Center (UTC) on Improving Rail Transportation Infrastructure Sustainability and Durability for financially supporting this research.

REFERENCES

- Alfredo, Fazio and J. L. Corbin. 1986. "Track Quality Index for High Speed Tracks." *Transportation Engineering* 1(112):46–61.
- Attoh-Okine, Nii. 2017. *Big Data and Differential Privacy*. Wiley Series in Operations Research and Management Science.
- Bartholomew, D. J. 2010. "Principal Component Analysis." in *International Encyclopedia of Education*.
- Bergmeir, Christoph, Gregorio Sáinz, Carlos Martínez Bertrand, and José Manuel Benítez. 2013. "A Study on the Use of Machine Learning Methods for Incidence Prediction in High-Speed Train Tracks." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7906 LNAI:674–83.
- Bishara, Anthony J. and James B. Hittner. 2017. "Confidence Intervals for Correlations When Data Are Not Normal." *Behavior Research Methods* 49(1):294–309.
- Brereton, Richard G. 2015. "The T-Distribution and Its Relationship to the Normal Distribution." *Journal of Chemometrics* 29(9):481–83.
- Buhmann, Joachim M., Gabriel Krummenacher, Seijin Kobayashi, Stefan Koller, and Cheng Soon Ong. 2017. "Wheel Defect Detection With Machine Learning." *IEEE Transactions on Intelligent Transportation Systems* 19(4):1176–87.
- Cárdenas-Gallo, Iván, Carlos A. Sarmiento, Gilberto A. Morales, Manuel A. Bolivar, and Raha Akhavan-Tabatabaei. 2017. "An Ensemble Classifier to Predict Track Geometry Degradation." *Reliability Engineering and System Safety* 161(April 2016):53–60.
- Charniak, Eugene. 2018. *Introduction to Deep Learning*. 1st ed. Cambridge, Massachusetts: The MIT Press.
- Cheng, Ruijun, Yongduan Song, Dewang Chen, and Long Chen. 2017. "Intelligent Localization of a High-Speed Train Using Lssvm and the Online Sparse Optimization Approach." *IEEE Transactions on Intelligent Transportation Systems* 18(8):2071–84.

- Dingqing, Li, James Hyslip, Ted Sussmann, and Steven Chrismer. 2016. *Railway Geotechnics*. Taylor & Francis.
- El-Sibaie, Magdy; and Yu-Jiang Zhang. 2014. "Objective Track Quality Indices." *Journal of the Transportation Research Board* 1863.
- Esveld, Coenraad. 2016. *Modern Railway Track, 2nd Edition*. MRT-Productions.
- FRA. 2002. "Track Safety Standards Compliance Manual." *Track and Rail and Infrastructure Integrity Compliance Manual* 4.
- FRA. 2018. *3.11 - Accident Detail Report*.
- Fumeo, Emanuele, Luca Oneto, and Davide Anguita. 2015. "Condition Based Maintenance in Railway Transportation Systems Based on Big Data Streaming Analysis." *Procedia - Procedia Computer Science* 53:437–46.
- Galván-Núñez, Silvia and Nii Attoh-Okine. 2018. "A Threshold-Regression Model for Track Geometry Degradation."
- Galván-Núñez, Silvia and Nii Attoh-Okine. 2018. "A Threshold-Regression Model for Track Geometry Degradation." *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 232(10):2456–65.
- Ghofrani, Faeze, Qing He, Rob M. P. Goverde, and Xiang Liu. 2018. "Recent Applications of Big Data Analytics in Railway Transportation Systems: A Survey." *Transportation Research Part C* 90:226–46.
- Hamid, A. and A. Gross. 1981. "Track-Quality Indices and Track-Degradation Models for Maintenance-of-Way Planning." *Transportation Research Record* (802):2–8.
- Han, Zhiwei, Alfredo Nunez, Zhigang Liu, Junwen Chen, and Hongrui Wang. 2017. "Automatic Defect Detection of Fasteners on the Catenary Support Device Using Deep Convolutional Neural Network." *IEEE Transactions on Instrumentation and Measurement* 67(2):257–69.
- Index, Quality and Quality Control. 2019. "Quality Indices for Managing Rail through Grinding Scholarly Articles for Quality Indices for Managing Rail through Grinding Use of Profile Indices for Quality

Control of Grinding - Arema (PDF) RCF Quantification - Towards Rail Grinding Best Practices F.”
5–6.

Jamshidi, Ali, Shahrzad Faghih-Roohi, Siamak Hajizadeh, Alfredo Núñez, Robert Babuska, Rolf Dollevoet, Zili Li, and Bart De Schutter. 2017. “A Big Data Analysis Approach for Rail Failure Risk Assessment.” *Risk Analysis* 37(8):1495–1507.

Krishnan, Vijaya. 2010. “Constructing an Area-Based Socioeconomic Index: A Principal Components Analysis Approach.” in *Early Child Development Mapping Project Alberta*.

Kullback, Solomon. 1978. *Information Theory and Statistics*. Peter Smith.

Lasisi, A., E. N. Martey, D. Guillot, and N. Atttoh-Okine. 2018. “A Three-Step Agglomerated Machine Learning: An Alternative to Weibull Defect Analysis of Rail Infrastructure.” in *Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018*.

Lasisi, Ahmed and Nii Atttoh-Okine. 2018. “Principal Components Analysis and Track Quality Index: A Machine Learning Approach.” *Transportation Research Part C: Emerging Technologies* 91(April):230–48.

Lasisi, Ahmed and Nii Atttoh-Okine. 2019a. “Machine Learning Ensembles and Rail Defects Prediction : A Multi-Layer Stacking Methodology.” *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*.

Lasisi, Ahmed and Nii Atttoh-Okine. 2019b. “Track Quality Index without Safety : Revisited .” *Transportation Infrastructure Geotechnology*, 7, pages1–12(2020).

Li, Hongfei, Dhaivat Parikh, Qing He, Buyue Qian, Zhiguo Li, Dongping Fang, and Arun Hampapur. 2014. “Improving Rail Network Velocity: A Machine Learning Approach to Predictive Maintenance.” *Transportation Research Part C: Emerging Technologies* 45:17–26.

Long, Zhiqiang, Lianchun Wang, and Ying Cai. 2009. “Comprehensive Fault Evaluation on Maglev Train Based on Ensemble Learning algorithm.” *Proceedings of the 2009 IEEE International Conference on Automation and Logistics, ICAL 2009* 2(August):1603–8.

- Lu, Shih Tong, Yao Chen Kuo, and Shih Heng Yu. 2010. "Risk Assessment Model for the Railway Reconstruction Project in Taiwan." *2010 International Conference on Machine Learning and Cybernetics, ICMLC 2010 2*(July):1017–22.
- Van Der Maaten, Laurens. 2009. "Fast Optimization for T-SNE." in *Proceedings of the Twelfth International Conference on Artificial Intelligence & Statistics*.
- Van Der Maaten, Laurens, Aaron Courville, Rob Fergus, and Christopher Manning. 2014. *Accelerating T-SNE Using Tree-Based Algorithms*. Vol. 15.
- Van Der Maaten, Laurens and Geoffrey Hinton. 2008. *Visualizing Data Using T-SNE*. Vol. 9.
- Martey, Emmanuel Nii, Lasisi Ahmed, and Nii Attah-Okine. 2017. "Track Geometry Big Data Analysis : A Machine Learning Approach." *2017 IEEE International Conference on Big Data (Big Data)* 3718–27.
- Mohammadi, Reza, Qing He, Faeze Ghofrani, Abhishek Pathak, and Amjad Aref. 2019. "Exploring the Impact of Foot-by-Foot Track Geometry on the Occurrence of Rail Defects." *Transportation Research Part C: Emerging Technologies* 102:153–72.
- Muinde, Michael Simba. 2018. *Railway Track Geometry Inspection Optimization*.
- Nielsen, Jens, Eric Berggren, Anders Hammar, Fredrik Jansson, and Rikard Bolmsvik. 2018. "Degradation of Railway Track Geometry – Correlation between Track Stiffness Gradient and Differential Settlement." *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 095440971881958.
- Oneto, Luca., Emanuele Fumeo., Giorgio Clerico., Renzo Canepa., Federico Papa., Carlo Dambra., Nadia Mazzino., and Davide Anguita. 2017. "Dynamic Delay Predictions for Large-Scale Railway Networks: Deep and Shallow Extreme Learning Machines Tuned via Thresholdout." *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 47(10):2754–67.
- Pedregosa, F., G. Varoquaux, and A. Gramfort. 2011. "Scikit-Learn: Machine Learning in {P}ython." *Journal of Machine Learning Research* 12:2825--2830.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, Thirion. B, Grisel. O, Blondel. M, Prettenhofer. P,

- Weiss. R, Dubourg. V, Vanderplas. J, Passos. A, Cournapeau. D, Brucher. M, Perrot. M, and Duchesnay. E. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12:2825--2830.
- Redfern, David and Douglas McLean. 2014. *Principal Component Analysis for Yield Curve Modelling Reproduction of Out-of-Sample Yield Curves*.
- Sadeghi, J. 2010. "Development of Railway Track Geometry Indexes Based on Statistical Distribution of Geometry Data." *Journal of Transportation Engineering* 136(8).
- Sadeghi, J., M. Najar, J. Zakeri, and C. Kuttelwascher. 2019. "Development of Railway Ballast Geometry Index Using Automated Measurement System | Elsevier Enhanced Reader." *Measurement* 138:132–42.
- Sadeghi, Javad, Hajar Heydari, and Elaheh Amiri Doloei. 2017. "Improvement of Railway Maintenance Approach by Developing a New Railway Condition Index." *Journal of Transportation Engineering* 146(3).
- Sadeghi, Javad, Amin Khajehdezfuly, ; Hajar Heydari, and Hossein Askarinejad. 2020. "Development of Railway Ride Comfort Prediction Model: Incorporating Track Geometry and Rolling Stock Conditions." *Journal of Transportation Engineering* 143(8).
- Sharma, Siddhartha, Yu Cui, Qing He, Reza Mohammadi, and Zhiguo Li. 2018. "Data-Driven Optimization of Railway Maintenance for Track Geometry." *Transportation Research Board Part C*.
- Soleimanmeigouni, Iman, Alireza Ahmadi, and Uday Kumar. 2018. "Track Geometry Degradation and Maintenance Modelling: A Review." *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 232(1):73–102.
- Su, Zhou, Ali Jamshidi, Alfredo Núñez, Simone Baldi, Bart De Schutter, A. Jamshidi@tudelft.nl, A. A. Nunezvicencio@tudelft.nl, and S. Baldi@tudelft.nl. 2019. *Integrated Condition-Based Track Maintenance Planning and Crew Scheduling of Railway Networks*.
- Su, Zhou and Bart De Schutter. 2018. "Optimal Scheduling of Track Maintenance Activities for Railway Networks." *IFAC-PapersOnLine* 51(9):386–91.

- Thaduri, Adithya, Diego Galar, and Uday Kumar. 2015. "Railway Assets: A Potential Domain for Big Data Analytics." *Procedia Computer Science* 53(1):457–67.
- Weston, P., C. Roberts, G. Yeo, and E. Stewart. 2015. "Perspectives on Railway Track Geometry Condition Monitoring from In-Service Railway Vehicles." *Vehicle System Dynamics* 53(7):1063–91.
- Wright, N. P., R. Gan, and C. McVae. 2016. "Software and Machine Learning Tools for Monitoring Railway Track Switch Performance." *7th IET Conference on Railway Condition Monitoring 2016 (RCM 2016)* 1–7.
- Yang, Chunsheng, Eric Trudel, and Yan Liu. 2017. "Machine Learning-Based Methods for Analyzing Grade Crossing Safety." *Cluster Computing* 20(2):1625–35.
- Zhang, Q. N., T. Huang, Y. X. Zhu, and M. K. Qiu. 2013. "A Case Study of Sensor Data Collection and Analysis in Smart City: Provenance in Smart Food Supply Chain." *International Journal of Distributed Sensor Networks* 2013.

List of Tables

1	Effect of Defect Predictors on Healthy Track Segment	29
D1	A summary of selected data science applications in Rail Track engineering	1
D2	Model Results with different Defect Predictors	2
D3	FRA Class 4 Threshold and Defect Summary	3

Draft

TABLE 1. Effect of Defect Predictors on Healthy Track Segment

Predictors	ALIGNMENT	GAGE and ALIGNMENT	GAGE	NO DEFECT	SURFACE
DBD	30	0	31	143372	0
TBD	8208	0	2	135166	57
Combined (DBD, TBD)	20055	0	19411	103875	92
3 PCs	7	0	0	143426	0
3D T-SNE	40304	0	0	103129	0

List of Figures

1	TQI, Hybrid Index and Track Maintenance	34
2	Potential Principal Component Threshold for Gage	35
3	Data Driven Approach to Maintenance	36
4	Selected Crucial TQI-based track geometry parameters.....	37
5	Data Structure	38
6	EDA: Correlogram and Feature-to-Feature Relationships.....	41
7	A simple MLP-Softmax Framework	44
8	Probability of Surface defect versus 1st Principal Component.....	45
D1	Cumulative Explained Variance.....	4
D2	Probability of Alignment defect Distribution	5
D3	EDA: Time and Distance-Based Track Geometry Parameters.....	6
D4	EDA: Boxplots	7
D5	EDA: Reduced Components	8

(a) Routine Maintenance with TQI

(b) Eliminating Corrective Maintenance with Hybrid Index

Fig. 1. TQI, Hybrid Index and Track Maintenance

Fig. 2. Potential Principal Component Threshold for Gage

Fig. 3. Data Driven Approach to Maintenance

Fig. 4. Selected Crucial TQI-based track geometry parameters.

Fig. 5. Data Structure

(a) Boxplots of Selected Geometry Parameters

(b) Boxplots of Selected Geometry Parameters

Fig. 6. EDA: Correlogram and Feature-to-Feature Relationships

Fig. 7. A simple MLP-SoftMax Framework

Fig. 8. Probability of Surface defect versus 1st Principal Component