1. Report No. RailTEAM UD-5	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Principal Components Analysis and Track Quality Index: a Machine Learning Approach		5. Report Date May 2019 6. Performing Organization Code:
7. Author(s) Ahmed Lasisi and Nii Attoh-Okine https://orcid.org/0000-0001-5328-5538		8. Performing Organization Report No. UD-5
9. Performing Organization Name and Address Department of Civil & Environmental Engineering University of Delaware 301 DuPont Hall Newark, DE 19716		10. Work Unit No. 11. Contract or Grant No. 69A3551747132
12. Sponsoring Agency Name and Address Office of Research, Development and Technology (RD&T) US Department of Transportation 1200 New Jersey Avenue, SE Washington, DC 20590		13. Type of Report and Period 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Track geometry data exhibits classical big data attributes: value, volume, velocity, veracity and variety. Track Quality Indices-TQI are used to obtain average-based assessment of track segments and schedule track maintenance. TQI is expressed in terms of track parameters like gage, cross level, etc. Though each of these parameters is objectively important but understanding what they collectively convey for a given track segment often becomes challenging. Several railways including passenger and freight have developed single indices that combines different track parameters to assess overall track quality. Some of these railways have selected certain parameters whilst dropping others. Using track geometry data from a sample mile track, we demonstrate how to combine track geometry parameters into a low dimensional form (TQI) that simplifies the track properties without losing much variability in the data. This led us to principal components. To validate the use of principal components as TQI, we employed a two-phase approach. First phase was to identify a classic machine learning technique that works well with track geometry data. The second step was to train the identified machine learning technique on the sample mile-track data using combined TQIs and principal components as defect predictors. The performance of the predictors was compared using true and false positive rates. The results show that three principal components were better at predicting defects and revealing salient characteristics in track geometry data than combined TQIs even though there were some correlations that are potentially useful for track maintenance.

17. Key Words Rail Infrastructure, Machine Learning, Track Quality Index, Data Science, Safety	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161. http://www.ntis.gov		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 21	22. Price

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

USDOT Tier 1
University Transportation Center
on Improving Rail Transportation
Infrastructure Sustainability and Durability

Final Report UD-5

PRINCIPAL COMPONENTS ANALYSIS AND TRACK QUALITY INDEX: A MACHINE LEARNING APPROACH

By

Ahmed Lasisi, Graduate Student
Department of Civil and Environmental Engineering
University of Delaware
Newark, DE

and

Nii Attoh-Okine, Ph.D., P.E., F. ASCE, Snr Member IEEE
Department of Civil and Environmental Engineering
University of Delaware
Newark, DE

Date: May 2019

Grant Number: 69A3551747132

DISCLAIMER

ABSTRACT

Track geometry data exhibits classical big data attributes: value, volume, velocity, veracity and variety. Track Quality Indices-TQI are used to obtain average-based assessment of track segments and schedule track maintenance. TQI is expressed in terms of track parameters like gage, cross level, etc. Though each of these parameters is objectively important but understanding what they collectively convey for a given track segment often becomes challenging. Several railways including passenger and freight have developed single indices that combines different track parameters to assess overall track quality. Some of these railways have selected certain parameters whilst dropping others. Using track geometry data from a sample mile track, we demonstrate how to combine track geometry parameters into a low dimensional form (TQI) that simplifies the track properties without losing much variability in the data. This led us to principal components. To validate the use of principal components as TQI, we employed a two-phase approach. First phase was to identify a classic machine learning technique that works well with track geometry data. The second step was to train the identified machine learning technique on the sample mile-track data using combined TQIs and principal components as defect predictors. The performance of the predictors was compared using true and false positive rates. The results show that three principal components were better at predicting defects and revealing salient characteristics in track geometry data than combined TQIs even though there were some correlations that are potentially useful for track maintenance.

Keywords: *Rail Infrastructure, Machine Learning, Track Quality Index, Data Science, Safety*

CONTENTS

DISCLAIMER ii
ABSTRACT iii
LIST OF FIGURES v
LIST OF TABLES vi
EXECUTIVE SUMMARY 1
INTRODUCTION 4
DATA PREPROCESSING 4
EXPLORATORY DATA 6
PRINCIPAL COMPONENTS FOR 150FT AND 500FT SECTIONS WITH SCREE PLOTS 8
PERCENTAGE AND CUMMULATIVE PERCENTAGE VARIANCE EXPLAINED 8
Defects and Defective Sections 9
Sections with Defects 10
CLASSIFICATION OF DEFECTIVE SECTIONS USING BEST MODEL AND 4-CLASS OF PREDICTORS INCLUDING 1ST 3 PRINCIPAL COMPONENTS 11
Biplots for A Sample Defective and Non-Defective Sections 12
CONCLUSIONS 13
ACKNOWLEDGEMENT 14
REFERENCES 14
ABOUT THE AUTHOR(S) 15

LIST OF FIGURES

Figure 1: Correlogram of Single, Artificial Indices and Principal Components ……....1
Figure 2: Track quality indices, tolerances and defects (Ciobanu 2016) ……...4
Figure 3: A sample parameter matrix with 35 sections (150ft) and 28 inspection data …….5
Figure 4: Track geometry parameters……..6
Figure 5a Processed data for 500ft section with row = inspection dates, column = parameters …7
Figure 5b: Pairwise scatter plot of Section 1 with 500ft section length……..8
Figure 6: Variance (LHS) and Cumulative Variance (RHS) explained by Principal Components for Sample Sections in 150ft (above) and 500ft (below) lengths……9
Figure 7 The principal component scores and the loading vectors in a single biplot display …..12
Figure 8 1st Two/Three Principal Components Plots for Both Defective and Non-Defective….12
Figure 9 SVM Classification on Two Principal Components Using a Radial Kernel ……..13

LIST OF TABLES

Table 1:Summary of Principal Components per Section 150ft and 500ft 1
Table 2: Summary of Principal Components for each Section ……..8
Table 3: FRA Safety Standards for Track Geometry Parameters ……10
Table 4: Defect Sections and Counts ……10
Table 5: Error Rates for Different Training Models ……..11
Table 6: Error Rates Using Different Training Parameters ……..11

EXECUTIVE SUMMARY

This study examines the potential of machine learning applications in railway track engineering. In this report, we investigate the possibility of reducing multivariate track geometry indices into a low-dimensional form without losing much information. Similar to the Pavement Condition Index in highways wherein weights are assigned to each parameter and then summed up (Karim et al. 2016).

However, author's proposed approach takes cognizance of the fact the observed multidimensional data often lies in an unknown subspace of two to three dimensions (Hastie et al. 2009). Hence, detecting this subspace in track geometry data can significantly enable authors to eliminate redundant information. This will make it possible to visualize multidimensional track geometry data in two or three dimensions which was hitherto impossible with the raw parameters obtained from track geometry cars. The second section of this report focuses on introducing track geometry parameters, data collection and track quality indices. The third section considers selected machine learning methods that are used to train, test and validate the use of single and combined track quality indices including the proposed principal components. Low-dimensional representation of multivariate track geometry parameters in terms of principal components was validated and compared to existing TQIs in the penultimate section. The last section of this report highlights key findings with concluding remarks.

This report formally described the work on principal components and track quality indices. To summarize heterogeneous track geometry data, some railways assign weights to selected track geometry parameter. This assignment is followed by the sum of all the products of the weights and the parameters to arrive at a value that is used as a measure of overall track quality. While the assigned weights are often subjective, the parameters selected vary from one railway to the other. Also, relevant information is lost through neglected parameters and subjective weight assignment. In order to prevent this, the use of principal components as combined TQIs was proposed in this work. This made it possible to simplify track geometry data in a way that most of the variance in the data is captured.

Table 3:Summary of Principal Components per Section 150ft and 500ft

	1^{ST} PC	**1^{ST} & 2^{ND} PCs**	**1^{ST}, 2^{ND} & 3^{RD} PCs**	**1^{ST}, 2^{ND}, 3^{RD} & 4^{TH} PCs**	**1^{ST}, 2^{ND}, 3^{RD}, 4^{TH} & 5^{TH} PCs**
Sections(150ft)	4, 5, 8, 15, 24, 25, 33	6, 9, 10, 11, 12, 13, 14, 17, 18, 26, 29, 30, 32, 35	1, 2, 7, 19, 20, 22, 23, 27, 28, 34	3, 16, 31	21
%(Count)	20(7)	**40**(14)	28.57(10)	8.57(3)	2.86(1)
Sections(500ft)	2, 4, 8, 10	3, 9, 11	1, 5, 6, 7	NA	NA
%(Count)	36.36(4)	27.27(3)	**36.36**(4)	0(0)	0(0)

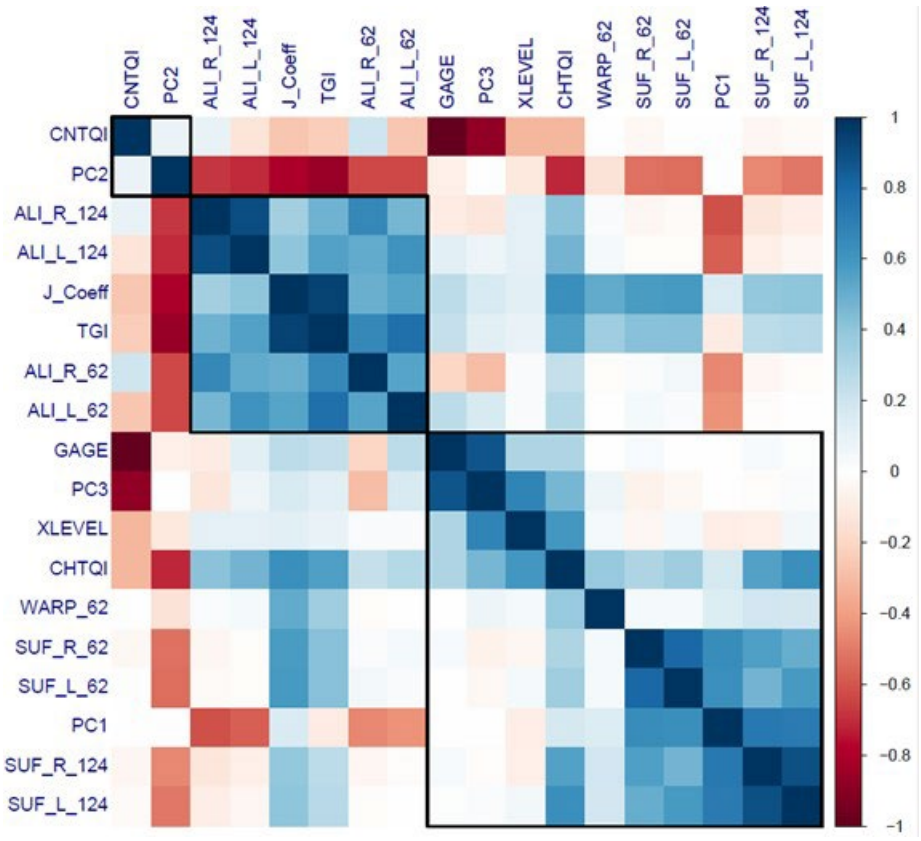

Figure 1: Correlogram of Single, Artificial Indices and Principal Components

The use of principal components as TQIs was tested using classical machine learning algorithms and the following conclusions are highlighted: (1) With a sample mile track, ⩾ 90% variance in the geometry data was explained by 1st 3 components in 100% of 500ft sections and 88% of 150ft sections. The first principal component captured track variations in the vertical plane, the second principal component in the longitudinal plane and the third correlated well with transverse irregularities. This information can be used to plan maintenance activities such as tamping or stone-blowing (1st PC and 2nd PC) as well as gage correction (3rd PC). (2) Support Vector Machine (SVM) was the most effective learning tool for classifying track sections with geometry defects among other selected machine learning tools; and (3) Using principal components and other combined TQIs from different railways, SVM predicted track defects better with 3 principal components and Canadian TQI than any other TQIs considered in this study. The prediction performance was measured using TPR (True Positive Rate) and FPR (False Positive Rate) since the defect data is highly unbalanced. This approach will help railways and track engineers assess track geometry monitoring from a different perspective as a novel method of combined/artificial TQI for maintenance scheduling. This work is a first step in incorporating dimension reduction in track geometry data analysis using classical techniques. Future work will consider the development of thresholds for principal components through correlation studies with vertical or lateral accelerations on train; and the use of classifier fusion to obtain better predictions.

Because dimension reduction/feature extraction with machine learning have not been widely adopted in track geometry data and analysis, there is great potential for optimized maintenance scheduling under this approach.

INTRODUCTION

Track geometry is a description of the track in terms of its longitudinal (alignment), transverse (gage) and vertical properties (surface/profile and cross level). Other track parameters combine these track irregularities in two-dimensions or more, e.g. vertical and longitudinal (warp/twist). Track quality index on the other hand is a quantitative representation of ride quality in an attempt to distinguish a good track from a 'bad' one. At this point, it is important to distinguish between track index, defects, irregularities and how they contribute to derailments. Firstly, tracks are laid to meet very stringent construction standards. Wear and tear as a result of track usage and tonnage results in deviations from construction standards. These deviations are often found in rails, track geometry, structure, etc. Since track parameters are often defined by a nominal value which is the characteristic of an ideal track.

Deviations from these nominal values develop into track irregularities (Ciobanu 2016). These irregularities grow gradually until it reaches an unacceptable limit (maintenance threshold) that requires intervention. Nominal values for a parameter beyond this limit defines a defect as seen in Fig. 1. Track geometry defects left to propagate is likely to lead to derailments as discussed in Section 1. To evaluate, assess and make decisions based on each parameter per unit length of track is almost practically impossible because it results in tremendous data-points and hypersensitivities in variations. Therefore, TQI is employed as an aggregate measure of a given track geometry parameter over a specific length of track. Standard deviation, mean, power spectral density (PSD), etc. are among the common average-based measures used as TQIs. Next, we discuss crucial track parameters and track quality indices expressed in terms of individual parameters.

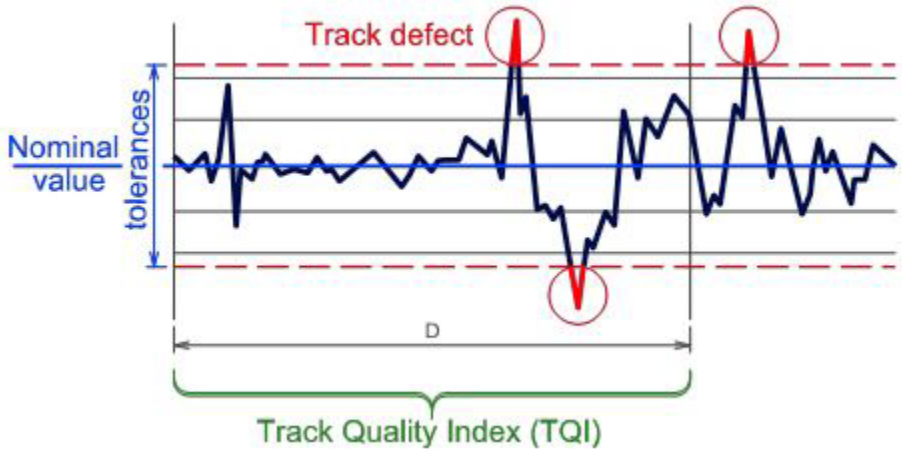

Figure 2: Track quality indices, tolerances and defects (Ciobanu 2016)

DATA PREPROCESSING

The dataset collected from a Class 7 track initially existed in a matrix format for each track parameter (e.g. Gage, Cross level, Alignment, Surface and Profile) in form of section lengths. We will be considering two section lengths only, 150ft and 500ft length. Other section length could be 62ft, 124ft, 200ft or even a 1000ft. The total length of track is about 5270ft which is equivalent to

about a mile. Therefore, the 150ft and 500ft section lengths are equivalent to 35 and 11 sections respectively. For a 500ft section, the Gage parameter matrix for instance is an 11 by 28 matrix where 11 stands for the number of sections and 28 represents number of inspection dates. Below Table 1 is an example of a typical parameter matrix represented as TQIs (standard deviation values) using the 150ft length.

Figure 3: A sample parameter matrix with 35 sections (150ft) and 28 inspection data.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	0.059799	0.061864	0.060197	0.058323	0.056592	0.059697	0.077568	0.077603	0.077886	0.07806	0.075433	0.055208	0.058491	0.063356
2	0.065605	0.038697	0.036298	0.039191	0.038058	0.039948	0.033053	0.033591	0.035318	0.033508	0.035762	0.037013	0.038903	0.037657
3	0.037132	0.025596	0.026045	0.026103	0.02616	0.027568	0.022861	0.0265	0.025851	0.023658	0.02678	0.030535	0.028031	0.032936
4	0.028333	0.043791	0.037435	0.04117	0.039777	0.04645	0.044381	0.043083	0.040056	0.036437	0.039464	0.034124	0.033915	0.041094
5	0.043572	0.045247	0.037243	0.041367	0.040747	0.043111	0.03102	0.030903	0.028835	0.028779	0.028948	0.026959	0.030111	0.031188
6	0.036907	0.044887	0.042565	0.041296	0.041735	0.039314	0.039963	0.039773	0.038719	0.039128	0.040101	0.035151	0.039631	0.040517
7	0.069351	0.051844	0.050604	0.051791	0.049985	0.049844	0.050638	0.054316	0.056625	0.059196	0.055023	0.045999	0.046916	0.054583
8	0.107111	0.115579	0.110836	0.110813	0.111417	0.113222	0.109803	0.129424	0.132856	0.133582	0.125291	0.110856	0.111529	0.114285
9	0.125718	0.124367	0.121614	0.120087	0.119355	0.124532	0.107449	0.11432	0.112555	0.112404	0.116055	0.042283	0.045032	0.050934
10	0.034126	0.054783	0.059263	0.053474	0.05698	0.060313	0.034987	0.040577	0.041761	0.041505	0.045244	0.034121	0.035526	0.038901
11	0.048397	0.030567	0.027014	0.026219	0.027414	0.026749	0.021664	0.026606	0.02223	0.022781	0.024184	0.025943	0.028027	0.027189
12	0.047304	0.028395	0.026788	0.028823	0.028318	0.02851	0.028887	0.033582	0.031593	0.029331	0.031746	0.04331	0.044614	0.043828
13	0.048688	0.046847	0.04969	0.052539	0.049891	0.051347	0.064216	0.067413	0.065692	0.069471	0.066609	0.059977	0.065483	0.070992
14	0.074558	0.082166	0.07747	0.082747	0.079994	0.082509	0.079185	0.086656	0.088578	0.090076	0.085381	0.045949	0.048065	0.051272
15	0.067064	0.066615	0.067526	0.066188	0.065973	0.062836	0.065489	0.076533	0.075342	0.072845	0.071817	0.038778	0.039248	0.038098
16	0.04863	0.036862	0.038397	0.0387	0.04011	0.039401	0.03233	0.037562	0.037784	0.040864	0.039377	0.032947	0.034798	0.034863
17	0.029184	0.035326	0.036558	0.034227	0.032675	0.036567	0.041009	0.04181	0.038722	0.039996	0.04093	0.029725	0.030879	0.03195
18	0.043823	0.039188	0.039919	0.039915	0.037348	0.040064	0.031002	0.033309	0.033614	0.033651	0.031235	0.049871	0.048757	0.046672
19	0.025728	0.026334	0.024067	0.02455	0.025835	0.022483	0.022672	0.027356	0.024472	0.025242	0.026064	0.025165	0.029331	0.025399
20	0.041181	0.034732	0.044058	0.042748	0.043851	0.044019	0.041464	0.049541	0.049873	0.050716	0.047819	0.056643	0.058278	0.053228
21	0.061551	0.033449	0.032277	0.033141	0.033971	0.032852	0.034573	0.033298	0.033063	0.033683	0.03443	0.035077	0.037492	0.034407
22	0.035573	0.034451	0.034683	0.036366	0.032957	0.035475	0.030588	0.036688	0.034285	0.034007	0.036342	0.03609	0.037481	0.038832
23	0.051892	0.056603	0.05544	0.055657	0.056222	0.055607	0.040662	0.043691	0.044359	0.043976	0.043783	0.044468	0.042389	0.04933
24	0.047649	0.038518	0.035664	0.035774	0.035232	0.035581	0.034896	0.034663	0.035348	0.037039	0.035045	0.037382	0.042858	0.038334
25	0.041239	0.028365	0.028145	0.025797	0.027729	0.026728	0.021663	0.024517	0.022691	0.022911	0.022283	0.027174	0.02708	0.02285
26	0.029667	0.03061	0.030554	0.031854	0.030875	0.029999	0.021623	0.025066	0.023539	0.02406	0.023443	0.023457	0.025749	0.022149
27	0.07116	0.071965	0.075571	0.071452	0.071186	0.073596	0.073363	0.073725	0.074785	0.072746	0.069639	0.073167	0.071029	0.07193
28	0.060196	0.035108	0.033476	0.037208	0.035475	0.035956	0.035474	0.0387	0.036966	0.035009	0.036453	0.035243	0.035523	0.034706
29	0.055968	0.083433	0.07293	0.077655	0.076462	0.075589	0.085294	0.076356	0.075817	0.076599	0.080788	0.081574	0.082029	0.082154
30	0.174825	0.110707	0.096511	0.09543	0.09824	0.1018	0.104779	0.097228	0.096645	0.096389	0.103789	0.105856	0.10253	0.103002
31	0.033837	0.04479	0.046112	0.047228	0.04436	0.052131	0.026907	0.036588	0.038194	0.035393	0.033409	0.03494	0.034795	0.03021
32	0.049662	0.054408	0.048215	0.047804	0.049509	0.05035	0.044831	0.04489	0.048469	0.04722	0.047765	0.052094	0.049919	0.051883

There are about 20 parameters collection from the field, 11 of these parameters have been selected relevant for this study. These parameters include: 1. Gage, 2. Cross level, 3. Surface Right (62ft), 4. Surface Right (124ft), 5. Surface Left (62ft), 6. Surface Left (124ft), 7. Alignment Right (62ft), 8. Alignment Right (124ft), 9. Alignment Left (62ft), 10. Alignment Left (124ft), and 11. Warp (62ft). Figure 4 shows the definition of these parameters.

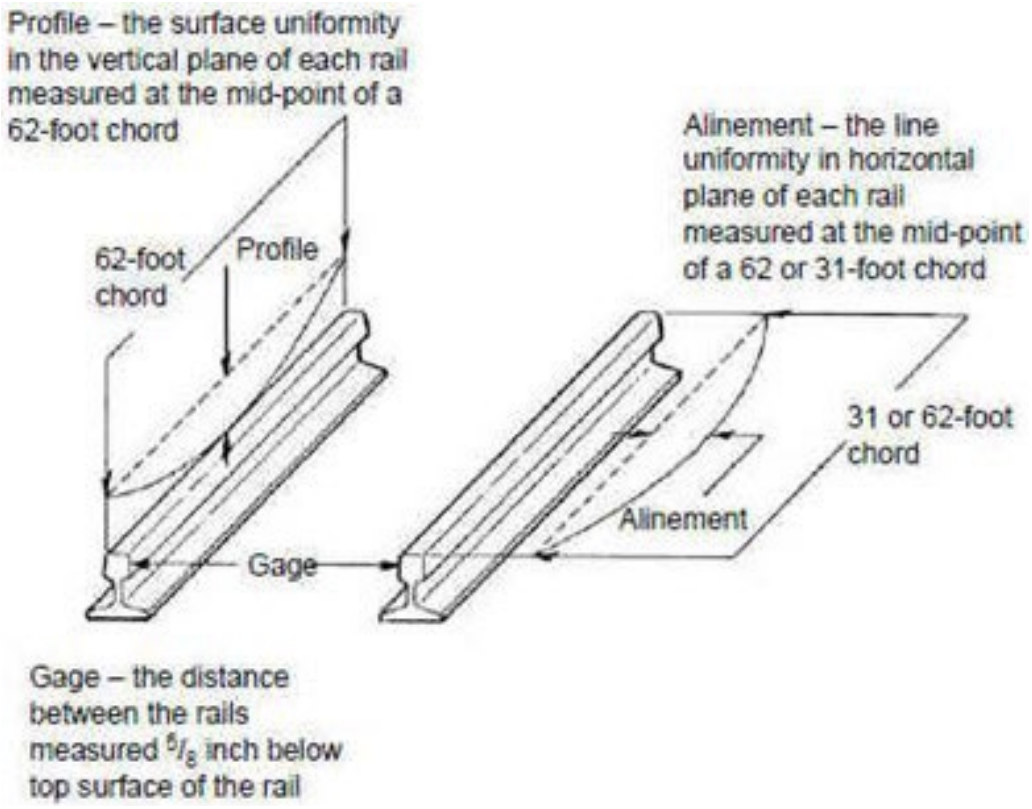

Figure 4: Track geometry parameters

EXPLORATORY DATA

Below Figure 5 is a sample of the processed data for the first section of the 500ft section length and scatter plot.

	Gage	Crosslevel	Surface Right (62 ft)	Surface Right (124 ft)	Surface Left (62 ft)	Surface Left (124 ft)
0613	0.03150941	0.06349561	0.06290028	0.1384306	0.05726861	0.12876667
0713	0.02895228	0.06786541	0.05890846	0.1337000	0.05519888	0.12655767
0813	0.02892818	0.05607177	0.05749043	0.1338793	0.05449526	0.12720271
0913	0.02842167	0.05361711	0.05772743	0.1343767	0.05396971	0.12497813
1013	0.02910333	0.05580554	0.05839406	0.1340208	0.05357174	0.12562701
1213	0.02843859	0.05441375	0.06185173	0.1351291	0.05521120	0.12190488
0114	0.03033620	0.05094548	0.05205959	0.1072029	0.04588575	0.09209619
0314	0.03055034	0.04998361	0.05313760	0.1092636	0.05014592	0.09488891
0414	0.02997477	0.05254619	0.05165735	0.1087922	0.04643317	0.09460935
0614	0.02944676	0.05189522	0.05023189	0.1172682	0.04682938	0.10476176
0714	0.02910582	0.04955953	0.05058808	0.1207248	0.04707749	0.10885900
1014	0.03519807	0.04827923	0.04040372	0.1062633	0.04724085	0.10123096
1114	0.03504658	0.05024769	0.04198325	0.1027350	0.04858260	0.09911142

Figure 5a Processed data for 500ft section with row = inspection dates, column = parameters

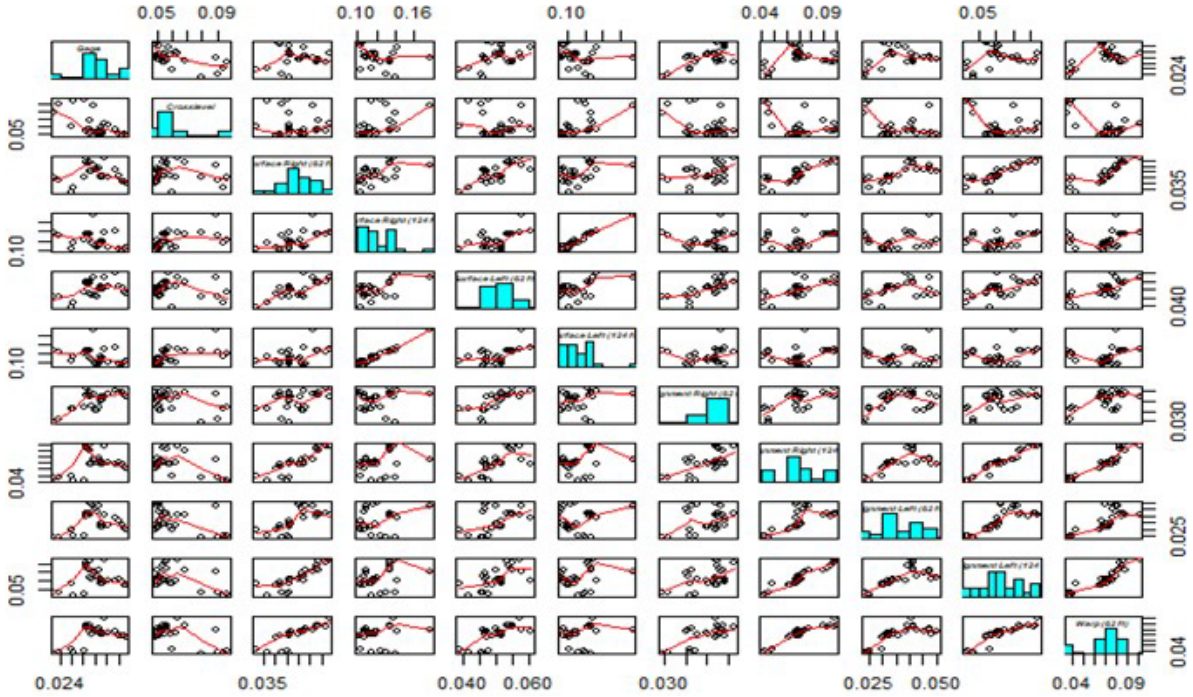

Figure 5b: Pairwise scatter plot of Section 1 with 500ft section length.

PRINCIPAL COMPONENTS FOR 150FT AND 500FT SECTIONS WITH SCREE PLOTS

Below is a list summary of the principal components that effectively summarize over 90% of the variation within parameters for each class of section length. Red inks denote sections summarized by only one principal component. Parameters are not scaled since they already exist as standard deviation with a general unit expressed in inches.

Table 4: Summary of Principal Components for each Section

Section Length = 150 feet, 35 Sections				Section Length = 500 feet, 11 Sections			
Section No.	1st Cumulative PCs > 90%	Percentage (%) explained	1st PC (%)	Section No.	1st Cumulative PCs > 90%	Percentage (%) explained	1st 2PCs (%)
1.	1ST, 2ND & 3RD	94.68	50.48	1	1ST & 2ND	91.50	91.50
2.	1ST, 2ND & 3RD	95.38	51.25				
3.	1ST, 2ND, 3RD & 4TH	94.34	41.49				
4.	1ST	92.74	92.74	2	1ST	95.60	99.10
5.	1ST	94.13	94.13				
6.	1ST & 2ND	96.67	86.60				
7.	1ST, 2ND & 3RD	92.11	63.21				
8.	1ST	94.00	94.10	3	1ST & 2ND	96.23	96.23
9.	1ST & 2ND	94.00	85.40				
10.	1ST & 2ND	94.37	67.20				
11.	1ST & 2ND	91.20	59.97	4	1ST	91.60	95.48
12.	1ST & 2ND	92.67	86.54				
13.	1ST & 2ND	92.10	80.24				
14.	1ST & 2ND	97.00	81.56	5	1ST, 2ND & 3RD	97.61	89.05
15.	1ST	91.79	91.79				
16.	1ST, 2ND, 3RD & 4TH	95.70	62.84				
17.	1ST & 2ND	92.45	81.06				
18.	1ST & 2ND	90.48	59.10	6	1ST, 2ND & 3RD	94.96	85.40
19.	1ST, 2ND & 3RD	93.97	69.25				
20.	1ST, 2ND & 3RD	95.49	55.17				
21.	1ST, 2ND, 3RD, 4TH & 5TH	93.56	54.30	7	1ST, 2ND & 3RD	96.14	88.57
22.	1ST, 2ND & 3RD	92.93	74.81				
23.	1ST, 2ND & 3RD	97.16	61.00				
24.	1ST	93.20	93.20	8	1ST & 2ND	94.38	94.38
25.	1ST	95.43	95.43				
26.	1ST & 2ND	95.55	81.55				
27.	1ST, 2ND & 3RD	97.14	74.74				
28.	1ST, 2ND & 3RD	94.20	71.56	9	1ST, 2ND & 3RD	98.00	86.28
29.	1ST & 2ND	97.55	87.61				
30.	1ST & 2ND	90.48	65.66				
31.	1ST, 2ND, 3RD & 4TH	96.84	50.73	10	1ST	91.30	97.33
32.	1ST & 2ND	96.21	85.91				
33.	1ST	91.70	91.70				
34.	1ST, 2ND & 3RD	92.90	52.99	11	1ST & 2ND	95.52	95.52
35.	1ST & 2ND	98.10	88.83				

PERCENTAGE AND CUMMULATIVE PERCENTAGE VARIANCE EXPLAINED

From the above, it is obvious that the first two principal components summarize at least 85% of

the data at any given section. Rather than express track geometry parameters as a function of 11 or more parameters, they could be effectively expressed as a bivariate data as has been shown above. A scree plot sample for sections in both 150 and 500ft section length also gives elbows at two principal components as shown below.

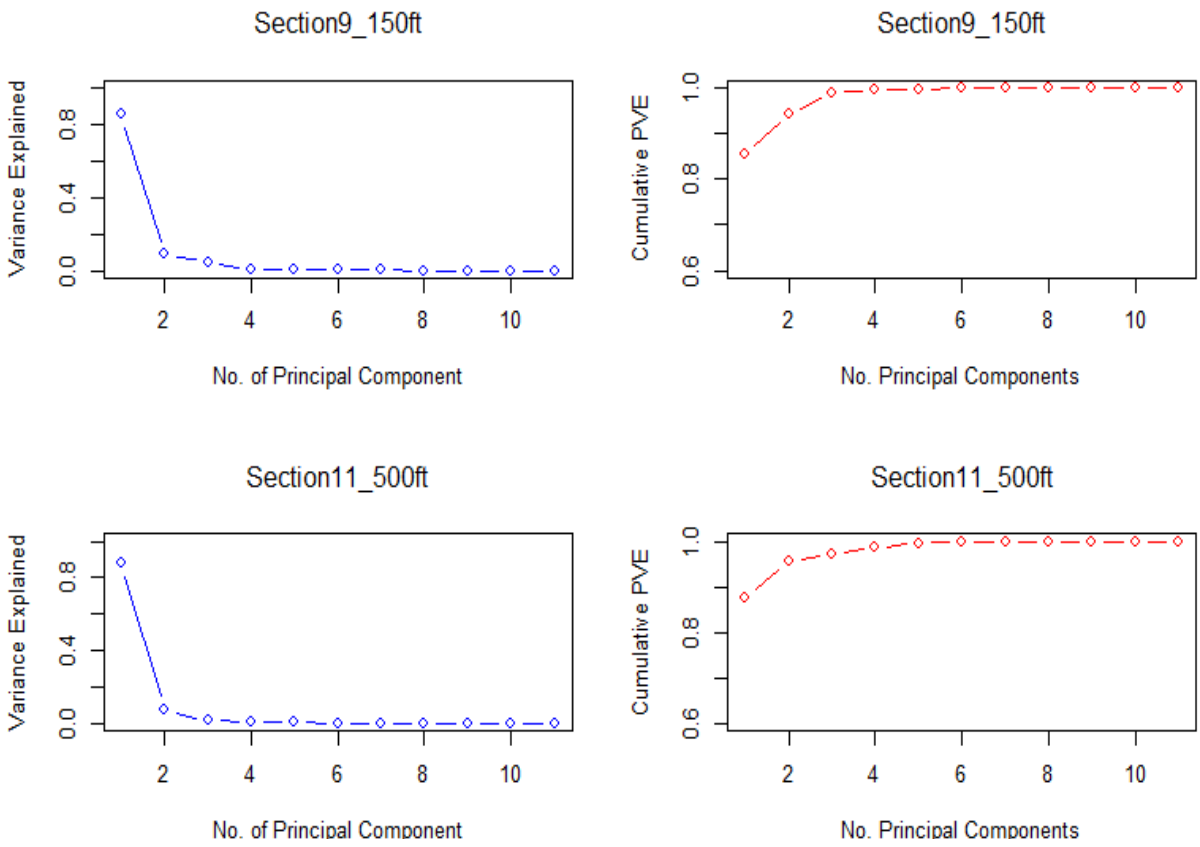

Figure 6: Variance (LHS) and Cumulative Variance (RHS) explained by Principal Components for Sample Sections in 150ft (above) and 500ft (below) lengths

Defects and Defective Sections

FRA safety standards: Below is a summarized table for the safety thresholds specified by the Federal Railroad Administration (FRA) for certain track geometry parameters relevant to this study. These thresholds are as follows:

Table 5: FRA Safety Standards for Track Geometry Parameters

Section #	Parameters for Class 7 Track	Safety Limits(inches)
1.0.	Gage	56''$\leq Gage \leq$ 57.25"
2.0.	Alinement 62ft	$\leq$ 0.5"
3.0.	Alinement 124ft	$\leq$ 1.25"
4.0.	Cross level	-0.5''$\leq Gage \leq$ 7"
5.0.	Surface 62ft	$\leq$ 1.0"
6.0.	Surface 124ft	$\leq$ 1.5"
7.0.	Warp 62ft	$\leq$ 1.5"

Sections with Defects

Firstly, a section with defect here is defined as the any point (in feet) within a section that violates at least one of the above thresholds as specified by FRA. This check was conducted for all sections across all inspection dates. Below is a summary of the sections with defects and their counts. Dates have not been included because this study is not concerned about degradation rate.

Table 6: Defect Sections and Counts

S/No	Parameters for Class 7 Track	Sections with Defects (500ft)	Location in Feet(s)	Count(s)	Total
1.0.	Gage	Section 6	2967 to 2971	5	5
2.0.	Alinement 62ft	No Defects	No Defects	0	
3.0.	Alinement 124ft	No Defects	No Defects	0	
4.0.	Cross level	Section 5 Section 5 Section 11 Section 11 Section 11	2056 to 2057 2108 to 2109 5194 to 5263 5233 to 5234 5232 to 5234	2 2 61 2 3	70
5.0.	Surface 62ft	No Defects	No Defects	0	
6.0.	Surface 124ft	No Defects	No Defects	0	
7.0.	Warp 62ft	No Defects	No Defects	0	
TOTAL					75

CLASSIFICATION MODELS AND ERROR RATES

The classification methods applied are three, two of which are parametric (Linear Discriminant Analysis and Support Vector Machine) and the other Non-parametric (Random Forest). All the defective sections were combined and these models were trained on them. Table 5 below shows the test/cross validation results for each of the models.

Table 7: Error Rates for Different Training Models

S/No	Learning Tool/Model	Training Error (%)	Test/CV Error (%)
1.	Linear Discriminant Analysis (LDA)	10.714	CV Error = 14.285
2.	Support Vector Machine (SVM)	8	Test Error = 5.8824
3.	Random Forest	0	Test Error = 5.88

Based on the above, it is interesting to see that the test error is actually lower than the training error rate with the SVM which is quite unusual and mostly the reverse often times. The test error on the Random Forest and SVM are however similar while the LDA is performing the least. The SVM will therefore be selected since it's parametric a bit more conservative to avoid overfitting.

CLASSIFICATION OF DEFECTIVE SECTIONS USING BEST MODEL AND 4-CLASS OF PREDICTORS INCLUDING 1ST 3 PRINCIPAL COMPONENTS

Table 8: Error Rates Using Different Training Parameters

S/No	Sections	All Parameters	J-Synthetic Coefficient	Indian TGI	1st 3 PCs
1.0.	Section 5	8.33%	8.33%	8.33%	0
2.0.	Section 11	10.714%	10.714%	10.714%	0

Biplots for A Sample Defective and Non-Defective Sections

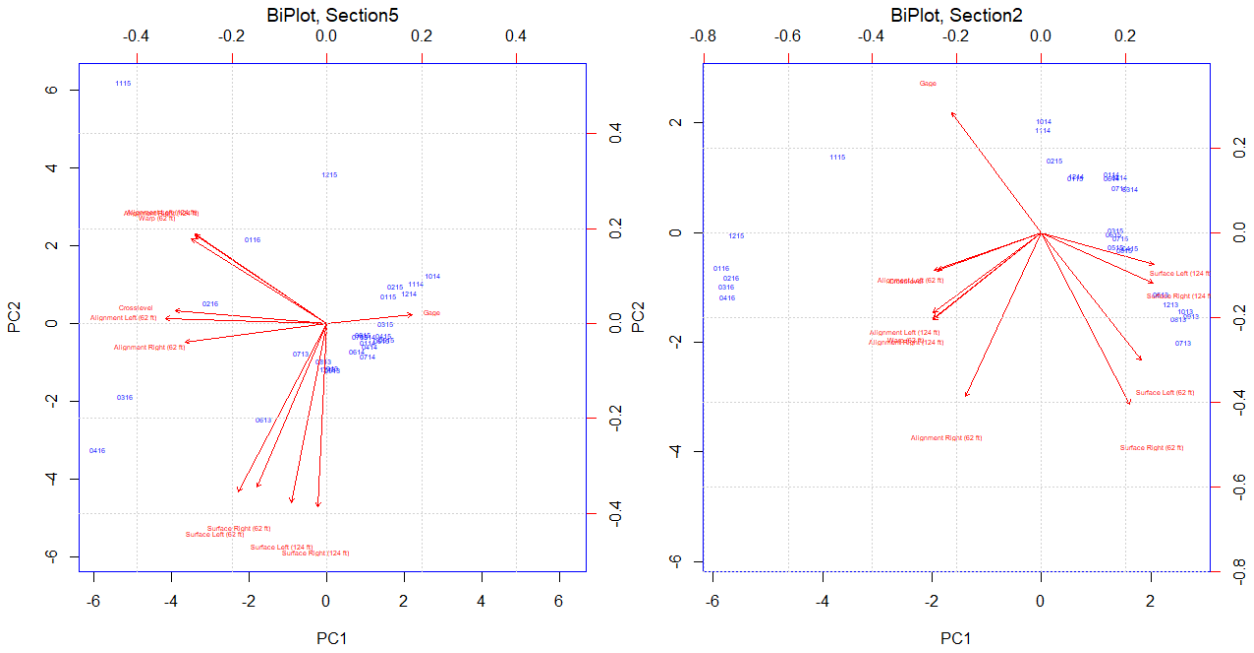

Figure 7 The principal component scores and the loading vectors in a single biplot display

The figures represent both the principal component scores and the loading vectors in a single *biplot* display. The axes with PC1 and PC2 are scores while the other two are the loadings for each component.

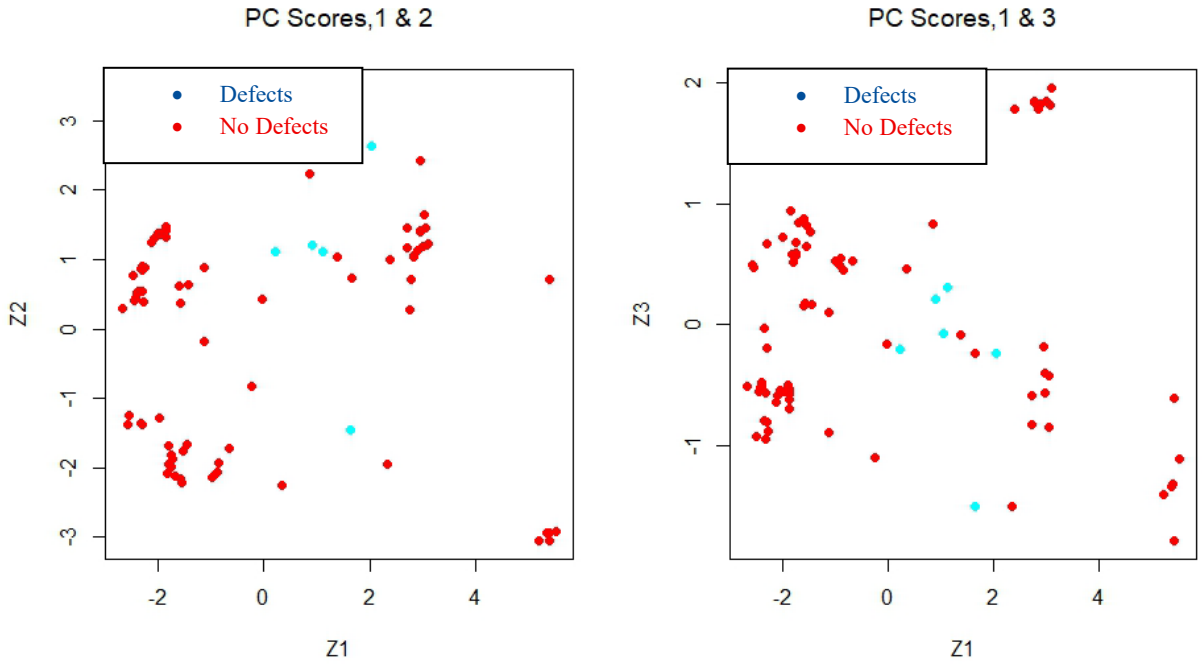

Figure 8 1st Two/Three Principal Components Plots for Both Defective and Non-Defective Sections

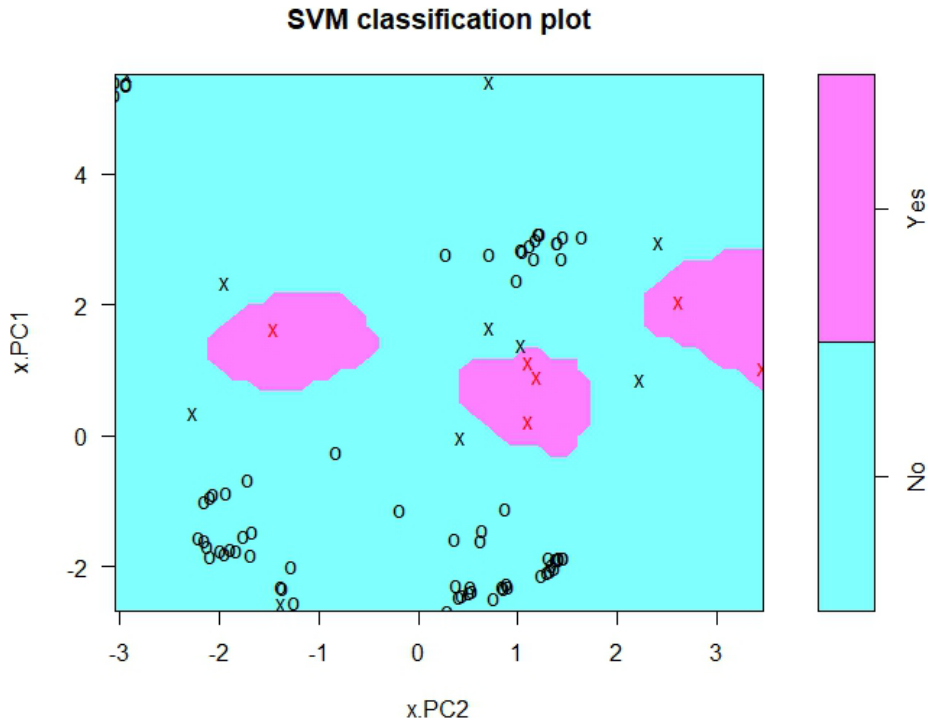

Figure 9 SVM Classification on Two Principal Components Using a Radial Kernel

CONCLUSIONS

This paper formally described the work on principal components and track quality indices. To summarize heterogeneous track geometry data, some railways assign weights to selected track geometry parameter. This assignment is followed by the sum of all the products of the weights and the parameters to arrive at a value that is used as a measure of overall track quality. While the assigned weights are often subjective, the parameters selected vary from one railway to the other. Also, relevant information is lost through neglected parameters and subjective weight assignment. In order to prevent this, the use of principal components as combined TQIs was proposed in this work. This made it possible to simplify track geometry data in a way that most of the variance in the data is captured.

ACKNOWLEDGEMENT

This study was conducted with the support from the USDOT Tier 1 University Transportation Center on Railroad Sustainability and Durability.

REFERENCES

1. Ciobanu, C., 2016. Evaluation of the track quality. *Pway Blog*. Available at: https://pwayblog.com/2016/09/11/evaluation-of-track-quality/?blogsub=confirming#blog_subscription-3.
2. Hastie, T., Tibshirani, R. & Friedman, J., 2009. *The Elements of Statistical Learning*, Available at: http://www.springerlink.com/index/10.1007/b94608.
3. Karim, F.M.A., Rubasi, K.A.H. & Saleh, A.A., 2016. The Road Pavement Condition Index (PCI) Evaluation and Maintenance: A Case Study of Yemen. *Organization, Technology and Management in Construction: an International Journal*, 8(1), pp.1446–1455.

ABOUT THE AUTHOR(S)

Ahmed Lasisi: Mr. Lassisi was a Research Assistant and Doctoral Candidate in the Railroad Program at the University of Delaware when this study was conducted. While he studied for his Ph.D. degree, his research interests include: Track Quality Safety, Spatial Analytics and Machine Learning Optimization, Transportation Infrastructure Improvements using multidisciplinary data analytical methods. Data-Driven Risk Analysis and Simulation Engineering. He obtained his Ph.D. in civil engineering from the University of Delaware, his Masters' and BS degrees from the University of Lagos of Nigeria in civil engineering.

.

Nii O. Attoh-Okine, Ph.D., P.E., F. ASCE, Snr Member IEEE

Nii O. Attoh-Okine, Professor of Civil and Environmental Engineering, and Electrical and Computer Engineering. He is also the Interim Academic Director of the University of Delaware Cybersecurity Initiative. In the last couple of years, he has authored two books which are defining the direction of research across disciplines: a) Resilience Engineering: Models and Analysis and b) Big Data and Differential Privacy in Railway Track Engineering. He is a founding associate editor for ASCE/ASME Journal of Risk and Uncertainty Analysis. He has served as an Associate Editor on the four ASCE Journals. Attoh-Okine is currently a member of a group of researchers from the United States and Japan working on Smart Cities and various cyber issues related to the Tokyo 2020 Olympic Games.