

Determining Sampling Points Using Railway Track Structure Data Analysis

Mikko Sauni¹[0000-0002-7034-5835], Heikki Luomala¹[0000-0002-7113-3527], Pauli Kolisoja¹[0000-0001-7709-180X] and Esko Turunen¹[0000-0002-1194-6497]

¹ Tampere University, Tampere 33720, Finland

mikko.sauni@tuni.fi

heikki.luomala@tuni.fi

pauli.kolisoja@tuni.fi

esko.turunen@tuni.fi

Abstract. In railway track asset management, limited funding is available to ensure safe and punctual train traffic on an aging rail network. Assessing railway structure problems, their severity and extent is a difficult and laborious task to which different methods have been applied. Further, determining problematic areas and identifying their rehabilitation needs are two separate operations.

In this research, data mining and data analysis of railway track structure data was used to identify different types of track behavior and corresponding substructure conditions. A descriptive data mining method, Generalized Unary Hypothesis Automata (GUHA), was adopted. Soil samples were taken and tested on the basis of the conducted analyses. The purpose was to see whether deductions made from the data, concerning the condition of the track substructure, could be confirmed with soil sampling and related soil sample laboratory tests.

The research was carried out in three parts. First, multiple data sources were used to comprise an initial data matrix, which was used in data mining and data analyses. After the analyses, fifty subballast and ten ballast sampling points were chosen according to the findings from data mining and data analysis, and samples were taken and tested. The last part of the research was to see how the laboratory test results corresponded with the analyses made from the data.

The research showed that GUHA data mining and data analysis can be used to detect sections of track with problematic substructures, but further research is required to improve the initial data.

Keywords: Data Analysis, Data Mining, GUHA Method, Railway, Sampling.

1 Introduction

Designing remediation for aged railway structures can be challenging, especially when using limited resources to do so. Design should begin with finding problematic areas, identifying different types of problems, and assessing remediation to those specific problems. Track geometry deterioration due to track structure degradation is observed at different rates in different sections of track. However, there is often uncertainty in explaining why some sections of track are more problematic than others.

Some problems of track structures may be more easily explained than others. Problems related to discontinuity in the rail or the track structure can be obvious, for example, problems with stiffness variations in bridge transitions. On the other hand, substructure problems can be difficult to detect because the problems might be invisible and dependent on seasons for example. For example, treating subgrade problems with just tamping will not fix the long-term problems of track geometry [1]. In this paper, track substructure and its components are defined according to Li et al. [2]; subballast is defined as the granular layer below the ballast and above the subgrade.

Plenty of data can be produced during the life cycle of a railway line including track geometry history, ground penetrating radar (GPR) measurements, track deflection measurements, and laser scanning point clouds, for example. However, the information about the track structure can be overwhelming to address manually.

Data mining and data analyses can be used to reveal interesting patterns in vast amounts of data. Data mining and data analyses have been used in the railway sector, for example to predict track geometry deterioration, optimize track maintenance, and model track asset management [3–6].

The Generalized Unary Hypothesis Automata (GUHA) data mining method can be used to describe an available data set. The method produces hypotheses based on the data, which are statements which the data either supports or does not support. The GUHA method can be used on historical data of the track structure to determine, for example, what has been the observed condition of the track structure on certain types of track structures.

In this research, the railway track structure condition of an old railway line in Finland was assessed using GUHA data mining and data analysis. Substructure material sampling points were determined according to data mining and data analyses results. The taken soil samples were tested for grain size distribution and capillary rise. The goal was to assess the applicability of the GUHA method to railway data and the validity of data currently available concerning railway substructures.

2 Background

2.1 Track Structure Degradation

In an ideal situation, newly constructed railway structures would settle uniformly, and the track geometry would remain undisturbed apart from descending. This situation cannot be achieved because the track structure nor the dynamic traffic loads on practically any given track section are homogenous and differences in settlements occur. Differences in track settlements cause the deterioration of track geometry on track sections and even in individual cross sections [7].

Shenton [8] has provided a comprehensive explanation for track geometry deterioration with six interacting factors: 1) dynamic forces 2) rail shape 3) sleeper spacing 4) sleeper support 5) ballast settlement 6) substructure differential settlement, in addition to all of the combinations and interactions of the aforementioned. Shenton has also provided causation, in which factors 2–6 increase dynamic loading if a train is moving

at a reasonable speed. Shenton's explanation is encompassing, but it does not explicitly elaborate what causes differential settlements in the substructure.

Li & Selig [9] have provided their explanation on the problems of earth structures, which consist of four factors and their interactions: 1) excessive loading (self-weight and repeated dynamic loading) 2) fine graded soils 3) high moisture 4) freezing and thawing.

Later, Li et al. [2] have identified the track substructure as the single most crucial element related to track performance. According to Li et al., a good track substructure has a strong resistance to plastic deformation and provides uniform elastic deformations, which can be achieved by having a well-drained layer consisting of angular and durable particles that resist abrasion. A poor substructure is described by a high moisture and high fines content [2].

According to these explanations, track structure degradation and its manifestation as track geometry deterioration can result from several reasons. However, if discontinuity in the rail or track structure and faults in the superstructure can be excluded with available information, problematic substructures should be detectable when fine graded soil materials or high moisture contents are observed.

2.2 GUHA Data Mining

Data mining, in general, is a collection of computer aided data analysis techniques that focus on discovering and modeling meaningful information in big data masses.

GUHA, an acronym for Generalized Unary Hypothesis Automata, is a logic-based method of a descriptive data mining procedure implemented to LISp-Miner software [10–12]. Given a large data matrix, in this research of 32 columns and 52,907 rows, acquiring answers to questions related to the data is possible. An encompassing description of the initial data used in this research is provided in section 3.2 and Table 2. Briefly summarized here, the columns in the data matrix represent properties (called predicates or attributes) the examined objects have, and each row characterizes an object; it either has the properties, does not have the properties, or the cell is empty. Examples of columns (predicates) in this research are structural thickness (12 categories with 0.2 m intervals), structural moisture index (7 categories), and data on assets such as bridges (3 categories) and culverts (binary categories).

Analytic questions are presented by using generalized quantifiers such as usually implies, above average, almost equivalent, etc. There are dozens of generalized quantifiers implemented in the LISp-Miner software. The software goes through the data where applicable and outputs dependencies that the data supports. Some of the answers may be already familiar to the user, but there may also be some interesting new ones. Next, three different LISp-Miner procedures used in this research are presented, as well as some interesting results found in the initial data.

Founded implication quantifiers are used to find dependencies: 'The presence of A implies the presence of B with confidence p and support n '. One of the questions presented in data mining was what kind of track structures exhibit a high track geometry deterioration rate with over 80% confidence when at least 700 rows of data support the statement. After 10,703,034 verifications, 238 hypotheses were supported by the data,

e.g., a hypothesis: a high track geometry deterioration rate is observed on 86% of structures that exhibit high moisture indices and high track deflection variance.

Above average quantifiers are used to find situations where among cases satisfying *A*, there are at least $100 * p\%$ more objects satisfying *B* than there are cases satisfying *B* in the whole data. This approach was used to get an answer to inquire, for example, which track structure properties or their combinations have an above average correlation to some type of track geometry deterioration rate. After 15,762,789 verifications, 278 hypothesis were supported by the data, e.g., a hypothesis: the observed track geometry deterioration rate is high on 96% of track structures that are less than 1.4 m thick, no frost insulation board are installed, and the track deflection mean and variance are high. The corresponding correlation is 17% on other structures.

In the Action mining approach, the idea is to find dynamic features in data; some predicates are considered stable attributes and some others are flexible attributes. One of the analytical questions examined in this research is the following: how do changes in the ballast moisture index influence the observed track geometry deterioration rate when other parameters are stable. After 715,806 verifications, 437 hypothesis were supported, e.g., when the track substructure thickness is 1.0–1.6 m and ballast thickness is < 500 mm, a low track geometry deterioration rate is observed on 75% of structures with a low ballast moisture and on 15 % of structures with a high ballast moisture.

The hypotheses are translated into comprehensible language from the contingency tables that the method and program produce. In practice, LISp-Miner tests 2x2-contingency tables obtained from the data matrix. For example, for the above average quantifier, to accept a dependence ‘*A* among those cases that satisfy *B* is 4 times more frequent than *A* in the whole data’, a condition

$$a/(a+b) > (1+p)((a+c)/(a+b+c+d)) \quad (1)$$

must be satisfied by the data in Table 1. For *A* ‘Track structures thickness < 1.4 m, no frost insulation board, track deflection mean is high, deflection variance is high’ and *B* ‘High track geometry deterioration rate’, the corresponding 2x2-contingency table is presented in Table 1; direct calculation shows that the data supports this dependence. Of course, every quantifier has its own condition for acceptance. The quantifiers and their formulas are presented in detail by Rauch [11].

Table 1. An example of a contingency table.

	B	not B
A	a = 501	b = 19
not A	c = 8,821	d = 43,466

3 Research Process

3.1 Case Track Section Kouvola–Kotka

The case study track section in this research, Kouvola–Kotka, is located in the south-eastern coastal area of Finland. The track section is a double track from Kouvola to

Juurikorpi and a single track from Juurikorpi to Kotka, with respective lengths of 35 km and 18 km.

The case study track section was originally completed as a single line track in 1890, and the double track sections have been built in the 1950s and 1990s. The age of the track section implies that the structures may not be fully compliant with today's standards. For example, the required minimum thickness of non-frost susceptible materials in a track structure on this area is 2 m, whereas the mean thickness of all structures is 1.96 m. This means that there are plenty of undersized structures along the track section. Further, the materials used may not be compliant with today's requirements.

Frost insulation boards are often installed into old track structures in Finland to reduce frost penetration into the track structure. On the case study track section, frost insulation boards have been installed on over 16 km of the 54 km track section. Frost insulation boards should be installed in the subballast 300 mm below the ballast layer, according to Finnish guidelines [13]. However, if frost insulation boards are installed when the ballast is undercut and cleaned, frost insulation boards are installed directly below the ballast layer. Unfortunately, documented information about the installation of frost insulation boards is rarely available.

3.2 Initial Data

The initial data included track geometry car measurements, GPR interpretations, track deflection measurements, laser scannings, and asset data (Table 2). The initial data was comprised into a single matrix, where the rows represented one-meter long sections of the track, which are described by their properties presented in the columns.

Table 2. Initial data sources, processing, and usage.

Data origin	Pre-processing	Data used for	Data type
Track geometry car measurements	Annual growth of running 20 m chord standard deviation	Track geometry deterioration rate	Ratio 1 variable
GPR	Signal rebound calculations	Structural layer thicknesses	Ratio 4 variables
GPR	Signal attenuation calculations	Moisture damage index MDI	Ratio 7 variables
Continuous laser scanning	Minimum elevation 2–6 m perpendicular from track center line	Ditch depth (both sides individually)	Ratio 2 variables
Continuous track deflection measurement	Running 20 m chord mean value	Track deflection	Ratio 1 variable
Continuous track deflection measurement	Running 20 m chord variance	Track deflection variation	Ratio 1 variable
Soil maps	Interpretation	Subsoil classification	Categorical 1 variable
Video and asset data warehouse	Visual inspection	Track assets (bridges, turnouts, culverts, etc.)	Categorical and binary 7 variables

The initial data was gathered in the same way over the entire length of the track section. On the double track section, the measurements and analyses concern the western track only. Continuous measurements included track geometry, GPR, laser scanning, and deflection measurements. Soil maps and asset data were homogenous too over the entire length of the track section. Video of the track was also provided, but it was used only to confirm and validate other data.

The initial data contained little missing values. The missing values were imputed using other available data. For example, an empty value for a bridge was set to mean that there was no bridge and video of the track was used to validate the imputation.

The input data for track geometry deterioration rate calculations was the track geometry car (Ttr1 51) semiannual measurements from 2008 to 2018. The track geometry deterioration rate was calculated using the annual growth of a 20 m chord standard deviation of the measurements. The mean of the track geometry deterioration rate for the whole track section was 0.14 mm/a. Over 75% of the track section displayed a track geometry deterioration rate less than the track section's average. This indicates that there are few sections of track where the track geometry deterioration rate is high, but on those sections, the rate is very high.

GPR measurements provided information on the substructural layer thicknesses and perceived moisture of the substructural layers. Layer thicknesses were used both independently and as a sum to indicate combined ballast, subballast, and embankment thickness. A moisture index was calculated separately for the ballast, subballast, and subgrade. Also, a combined value of the aforementioned, as described by Arnold et al. [14], representing the moisture content of the whole substructure and moisture damage index (MDI), was provided.

A laser scanning point cloud was used to calculate the ditch depth, in order to provide information about drainage conditions. The minimum depth 2–6 m perpendicular to the track center line was calculated. Further, a 20 m chord minimum of the aforementioned was calculated to reduce error due to foliage and wayside equipment. This value depicted the ditch depth and was calculated individually for both sides of the track.

A continuous track deflection measurement car, as presented in more detail by Luomala et al. [15], was used to measure track deflection. A 20 m chord mean and variance values were both used: the former to indicate the level of deflection and the latter to indicate changes in the deflection.

Other sources, such as soil maps, asset management data warehouses, and a video check were used to identify track assets that influence the performance of track structures. The track asset data included bridges, culverts, turnouts, level crossings, frost insulation boards, stations, cuttings, and subsoil assessments. This data was either binary or conformed to classes using dummy values such as 0 for embankment, 1 for rock cutting, and 2 for soil cutting.

3.3 Applying GUHA Data Mining

The initial data was used as an input in the LISp-Miner program, which is an application of the GUHA method. Questions about the correlations between track structure properties and the track geometry deterioration rate were inquired. Both outcomes, high and

low track geometry deterioration rates, were inquired from several viewpoints. Quantifiers p-implication (PIM) and above average dependence were used in modules 4ft-Miner, SD4ft-Miner, and Ac4ft-Miner, but only relevant queries were reported. The data backing the hypotheses was visualized, and video from the areas was checked to verify that no other explaining features outside the hypotheses' data could be observed.

All of the queries used track structure variables as antecedents and the track geometry deterioration rate as the succedent. Conditions were applied in query 5 and 6 to eliminate unwanted variables from data mining. Several other queries were also conducted and many hypotheses per query were generated, but only the relevant queries and relevant hypotheses to this research are presented. Altogether, 69 queries were made and thousands of hypotheses were generated.

The following queries about the data were conducted and reported. Detailed information about formation and outcomes the queries is provided in Table 3.

1. What kind of combination of track structure variables is associated with a certain type of track geometry deterioration rate with more than 90% confidence?
2. What kind of combination of track structure variables is associated with a higher than average track geometry deterioration rate with more than 80% confidence?
3. How does a change in the variable for structure moisture affect a certain type of track geometry deterioration rate when other track structure variables are stable?
4. How does a change in the variable for overall structure thickness affect the most common type of track geometry deterioration rate when other track structure variables are stable?
5. How does a change in the variable for overall structure thickness affect a certain type of track geometry deterioration rate when other track structure variables are stable, and structures founded only on embankments or soil cuttings are examined?
6. How does having a frost insulation board affect a certain type of track geometry deterioration rate when only structures that are 1.6 m to 2.4 m thick are examined?

The relevant non-trivial hypotheses are presented below. The hypothesis number corresponds to the query number, for example, hypothesis number one is a result of query number one.

1. Track geometry deterioration rate is low on 92% of structures that are built on embankments, over 2.8 m thick, and have over 650 mm thick ballast layers, low moisture indices, and little track deflection variance.
2. Track geometry deterioration rate is high on 87% of structures that are built on embankments, have a low embankment thickness (< 0.5 m), and the track deflection mean and variance are high.
3. Track geometry deterioration rate is low on 86% of structures with low moisture indices, whereas the corresponding percentage is 38% on structures with high moisture indices, when in both cases structures are built on embankments and do not have a frost insulation board in their structure.

4. A high track geometry deterioration rate is as common (about 84% of structures) on structures less than 1.4 m thick as a low track geometry deterioration rate is on structures 1.6–2.0 m thick, when the track is built on an embankment, there is no frost insulation board in the track structure, and the ballast layer is less than 450 mm thick.
5. Track geometry deterioration rate is low on 74% of structures less than 1.4 m thick, whereas the corresponding percentage is 36% on structures 1.8–2.4 m thick, when in both cases structures have a frost insulation board in their structure, and the track deflection mean and variance are low, and only structures founded on soil cuttings or embankments are regarded.
6. On structures that have deep ditches, track deflection is low, and a less than 550 mm thick ballast layer, the track geometry deterioration rate is low on 79% of structures that do not have a frost insulation board, whereas the corresponding correlation is 16% on equivalent structures that have a frost insulation board, when structures only 1.6–2.4 m thick are regarded.

Table 3. List of the reported LISp-Miner data mining tasks.

Query/ hypothesis	Module	Statistical quantifier	Frequencies quan- tifier	Verifica- tions	Number of hypotheses
#1	4ft	PIM > 0.9	a > 1,500	3,934,998	45
#2	4ft	PIM > 0.8	a > 700	10,703,034	238
#3	Ac4ft	PIM before > 0.7 & PIM difference > 0.4	a (before) > 2,000	134,344	48
#4	Ac4ft	PIM before > 0.8 & PIM after > 0.8	a (before and after) > 400	69,071,100	2
#5	Ac4ft	PIM difference > 0.3	a (before) > 1,000	45,736	25
#6	Ac4ft	PIM difference > 0.4	a (before) > 500	19,926	233

The hypotheses indicated that the moisture content, deflection variance, the thickness of track structure, and frost insulation boards affect the perceived performance of the track structure. The higher moisture content the structure displayed, the more track deflection the structure exhibited, and the thinner the structure was, the higher the track geometry deterioration rate was. These results are in line with the literature review in section 2.1.

Frost insulation boards produced somewhat conflicting results. Structures less than 1.4 m thick have performed well when a frost insulation board was detected in the structure. However, the data indicated poor performance on structures which were 1.6–2.4 m thick and had a frost insulation board in the structure. These structures would not generally require a frost insulation board if the materials used to build the subballast are not frost susceptible.

3.4 Data Analysis

The initial data was visualized in the Rail Doctor® program so that interesting structures could be located and sampling points planned. Using the information gained from literature (section 2.1) and data mining (section 3.3), structures with above average track geometry deterioration rates, structures with high moisture indices, thick structures with frost insulation boards, and structures with high deflection variation were targeted. The sampling points were manually selected using the above-mentioned criteria and avoiding discontinuity areas.

An example of data visualization is presented in Fig. 1. The x-axis represents track kilometers, and track structure data is represented on the y-axis. Starting from the top, the y-axis contains structural layer boundary depths, asset data, MDI in color maps, relative structural moisture in graphs, subgrade soil classification assessment, the track geometry deterioration rate, and track deflection.

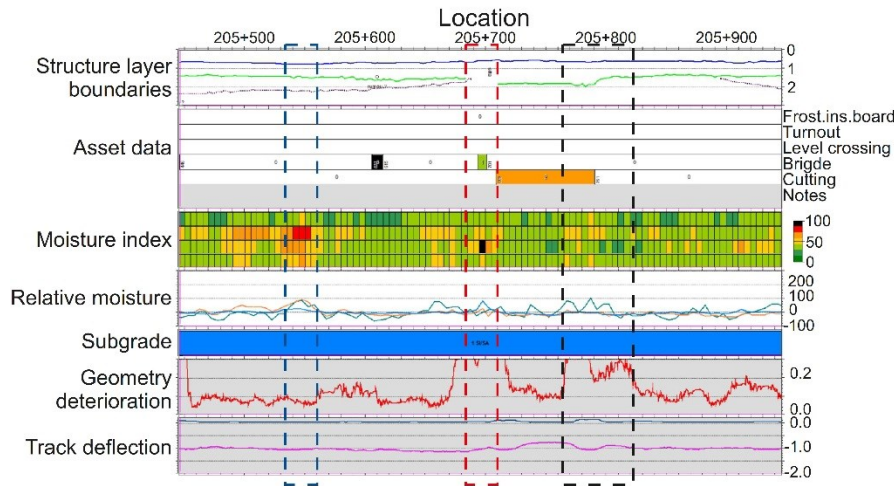


Fig. 1. Visualization of data in Rail Doctor®.

In Fig. 1, three areas are bordered by dashed lines. The left most bordered area exhibits high subballast and subgrade moisture on an over 2 m thick embankment. On that area, the track geometry deterioration rate is lower than average, and no track deflection variations are detected. The right most bordered area contains a soil cutting, where slightly moist subballast is observed, the track geometry deterioration rate is high, and track deflection is locally high, which is indicated in the mean and variance of track deflection. These two bordered areas are of interest with regard to subballast sampling.

Discontinuity in the rail or the track structure, such as bridges, turnouts, and culverts, could easily be detected when the data was visualized. Some of these areas were interpreted to be problematic due to their geometry history and deflection, which differed vastly from other sections of the track. An example of a problematic bridge transition can be found in the dashed area in the middle of Fig. 1. Two peaks in the track geometry

deterioration rate are observed in the transitions to the bridge, and track deflection fluctuates. Track discontinuity areas, while interesting, were not the subject of this research.

3.5 Sampling and Laboratory Tests

Fifty subballast and ten ballast sampling points were selected along with 15 and 2 back-up points for subballast and ballast, respectively. Various substructural conditions were required, therefore the subballast sampling points were selected as follows: 10 well performing points with a low moisture content, 10 well performing points with a high moisture content, and 30 problematic areas possibly due to substructure conditions. The ballast samples were selected from structures that were interpreted to be problematic according to the initial data.

The subballast samples were taken from two depths: 300–600 mm and 600–900 mm below the track bench (adjacent to the ballast shoulder). Some additional samples were taken from varying depths if a clear soil layer boundary was detected while taking the samples. Altogether, 118 subballast samples were taken.

All subballast samples were subjected to sieving in accordance with SFS-EN 933-1:2012 [16]. The grain size distribution and natural water content of all samples were investigated. Also, the coefficients of uniformity C_u and curvature C_c were calculated.

The grain size distributions were surprisingly similar throughout the track section. There were 33 samples in which the coefficient of uniformity C_u was less than 5, meaning that the range of particle sizes was narrow. Thirty-one samples exceeded the limit values given in Finnish guidelines for subballast materials on the fine graded side of the grain size distribution scale. Nevertheless, only six samples exceeded the fines (≤ 0.063 mm particle size) content limit of 4%. In some samples, a clear presence of the ballast material was detectable, but this was an expected result because the same observation was already determined while taking the samples.

The capillary rise test was subjected to 60 subballast samples which were taken mainly from the lower depth. Capillary rise was tested by placing samples in plastic tubes in shallow water for one week and measuring the highest visible waterline in the sample. This value represented the sample capillary rise. The average capillary rise of all samples was 31.5 cm. 5.5 cm and 66 cm were the minimum and maximum values, respectively. All samples exhibited very low fully saturated zones.

The ballast samples were retrieved and sieved according to a Finnish national guideline for ballast material sampling. The ballast sample was retrieved 30–40 cm below the bottom of the rail. A sample weighing 6–8 kg was taken, using a shovel, from between two sleeper ends from an area the size of 20 cm by 20 cm and 10 cm deep. Ballast samples were sieved, and material found on 1 mm, 8 mm, and 25 mm sieves was recorded. A ballast fouling index, described by the sum of the percent finer by weight on each of the aforementioned sieves, was used in determining sample quality. According to Finnish guidelines, if the ballast fouling index exceeds the limit value of 90, the ballast layer must be renewed or cleaned [17]. All ballast samples exhibited low ballast fouling indices, which were between 7.2 and 21.9. This means that all tested ballast materials are, according to the guideline, in very good condition.

4 Results and Discussion

4.1 Soil Sample Laboratory Test Result Analysis

Correlations between the soil sample laboratory test results and the initial data were investigated. Obtained relevant results are presented in this section. Fig. 2 and Fig. 3 show box plots of different property classes' track geometry deterioration rate of all samples and samples taken only from structures that do not have frost insulation boards. Fig. 4 presents a box plot of different property classes' d50 grain size concerning all samples. The boxes in the boxplot represents the second quartile of the data. The vertical lines (whiskers) represent the lowest and highest data points within the 1.5 inter quartile range of the lowest or highest quartile, respectively. Outlier points can be found outside the whiskers as empty dots. The crosses in the box plot are means and the horizontal lines are medians.

The properties in the graphs include capillary rise, frost insulation boards, MDI, d50 grain size, sample depth, material consistent with guideline grain size distribution for subballast (pass or fail), and coefficient of uniformity. Samples from structures with frost insulation boards were ignored in MDI boxes, because the GPR results beneath frost insulation boards are disrupted by the boards and do not represent true values.

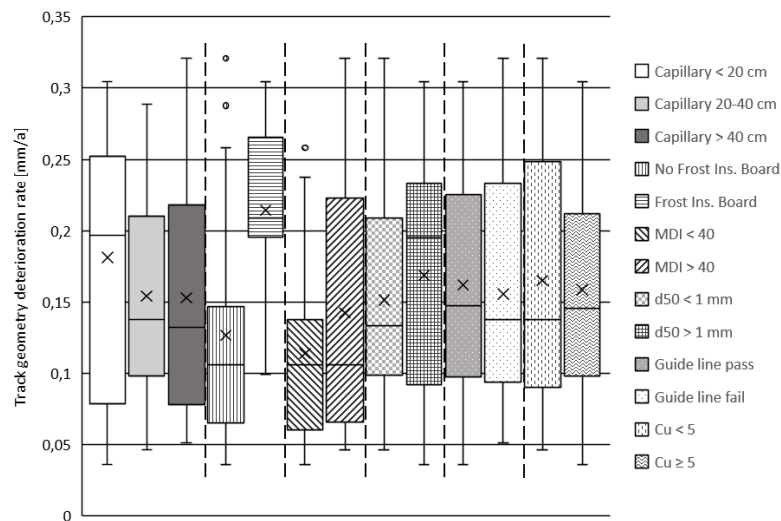


Fig. 2. Box plot of the track geometry deterioration rate of all samples.

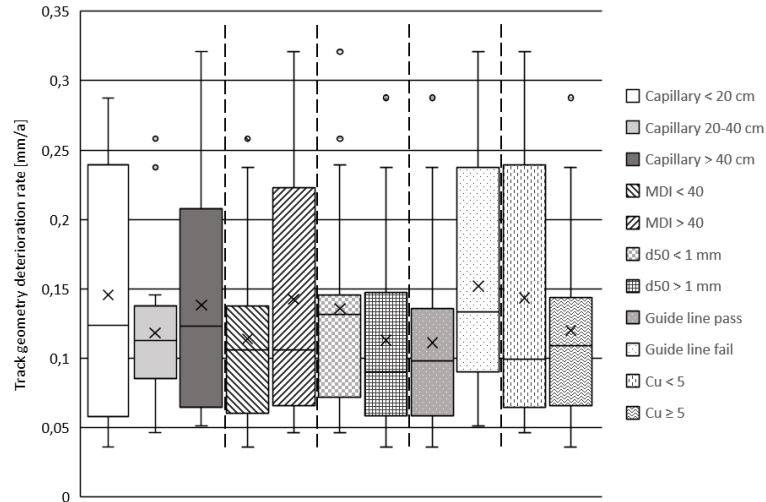


Fig. 3. Box plot of the track geometry deterioration rate of samples taken from structures that have no frost insulation boards.

Fig. 2 denotes that material properties had no practical effect on the track geometry deterioration rate if all samples are examined. Notable differences in the track geometry deterioration rate are observed only with regard to the MDI and frost insulation boards. High MDI and frost insulation boards are associated with a high track geometry deterioration rate. The effects of frost insulation boards, however, appear to be more dominant to the track geometry deterioration rate than subballast material properties.

If structures that have frost insulation boards are ignored, the influence of subballast material quality in the track geometry deterioration rate can be observed. In Fig. 3, the most distinct result in the sample material quality is that samples consistent with grain size distribution guidelines clearly exhibit lower track geometry deterioration rates compared to samples inconsistent with guidelines. The result, though, is not as notable as the effects of frost insulation boards in Fig. 2.

Further, in Fig. 3, a higher MDI, a smaller d50 grain size, and higher coefficient of uniformity values appear to have some correlation to higher track geometry deterioration rates. Unexpectedly, a low capillary rise appears to have a slightly higher track geometry deterioration rate compared with a high capillary rise.

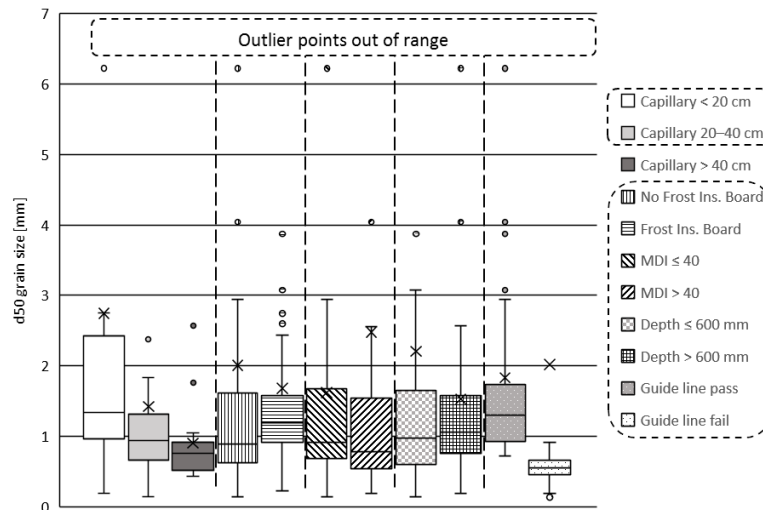


Fig. 4. Box plot of d50 grain size of all samples. Outlier points are outside of the graph range in all categories except ‘Capillary > 40 cm’.

Fig. 4 represents the distributions of d50 grain size with regard to other parameters. Outlier points are present even out of the plotting range (14–35 mm) due to ballast material found in the subballast material, which increases the mean values in all bars except the > 40 cm capillary rise. d50 grain size and capillary rise correlated intuitively; A lower capillary rise was detected on samples that had larger size particles and vice versa. Structures with or without frost insulation boards did not exhibit much variation in their d50 grain size distributions. Higher MDI values had a higher mean d50 than lower MDI values, even though the medians and upper and lower quartiles were similar. Samples taken from a lower depth had slightly smaller d50 grain sizes, but again, the medians and upper and lower quartiles were quite uniform. Material in accordance with design guidelines was clearly coarser than material that did not meet the guideline limits.

In many cases, though, GPR interpretations can indicate a significantly moist subballast layer, yet no major variation in the subballast material moisture content or quality was detected in the samples taken from these locations, even when structures with frost insulation boards are ignored. This may indicate that GPR interpretations are influenced by the prevailing conditions of the track, which are dependent on many more factors than just the subballast material. Another factor influencing the correlation between GPR interpretations and subballast sample test results is the representativeness of samples taken from the slope of the embankment in relation to the material directly beneath the center line of the track.

4.2 Findings from Sample Locations

The GUHA data mining indicated that frost insulation boards were associated with problematic structures when installed in thick structures. However, the materials found

in structures having frost insulation boards did not differ much from the non-problematic structures without frost insulation boards.

For instance, samples taken from around the 218-kilometer pole had no practical difference in their grain size distribution, natural water content, or capillary rise. However, the measured track geometry had deteriorated more than twice as fast on the section with frost insulation boards than that on the whole track section on average. As track structures or sub soil conditions do not vary much on this section, the only variable in the track structure is the frost insulation board.

Sampling revealed an extruded polystyrene foam frost insulation board installed directly underneath the ballast. The frost insulation board itself was in good condition, but the structure above and below the board was moist. Further, the board was covered with a fine graded material, which was most likely fouled ballast (Fig. 5).



Fig. 5. A frost insulation board underneath ballast at sample point 216+068. Photo credit: Toni Saarikoski.

The results indicate that more attention should be paid to the installation of frost insulation boards. The drawbacks of installing a frost insulation board directly under the ballast layer instead of installing it in the subballast should be investigated further because detrimental effects regarding structures' long-term performance may occur. Frost insulation boards may have a fouling effect on the ballast material due to increased stiffness variations in addition to drainage conditions that may be compromised.

Further, the ballast material of the sample point 216+068 was tested in accordance with guidelines in effect today in Finland. According to the results, the ballast material is of very good quality even though fouling is clearly detectable as presented in Fig. 5. This controversy results from taking the sample too high up in the ballast layer where a little amount of the fouled material can be found. These results give a good reason to review the guidelines in effect in Finland concerning ballast sampling and testing.

5 Conclusions

GUHA data mining and data analyses were used to assess the railway substructure conditions. Subballast and ballast sampling points were chosen according to the data mining and data analyses results. The material properties of the subballast samples and the initial data were used to examine substructural conditions and their effects on the track geometry deterioration rate.

The following conclusions were made in this research:

- The GUHA method is a novel approach to analyzing railway data. The practical benefits of using GUHA in analyzing railway data are best obtained in the early stages of designing maintenance or rehabilitation of an old railway structure, when maintenance data is abundant.
- The GUHA method could point out specific types of substructures to be problematic. The results from the GUHA data mining were in line with literature concerning explanations for track structure degradation.
- Using the knowledge gained from the GUHA method, the sampling can be focused on problematic structures, which would increase efficiency in design.
- The analyzed sampling data combined with the initial data indicated that frost insulation boards displayed a dominant correlation to the track geometry deterioration rate; structures with frost insulation boards were found to be more problematic than structures without frost insulation boards, even when the subballast material or substructure formation did not differ.
- However, if structures with frost insulation boards were ignored, material quality did exhibit correlations to the track geometry deterioration rate. In that case, a distinct correlation was detected between the subballast grain size distribution and track geometry deterioration rate. If the subballast sample grain size distribution was found to be consistent with current guidelines, a lower track geometry deterioration rate was observed.
- Individual material properties also resulted in an effect; coarse and well graded materials displayed a minor correlation to low track geometry deterioration rates. Further, a high MDI interpreted from GPR measurements displayed a considerable correlation to a high track geometry deterioration rate.
- GPR interpretations of soil moisture content did not always correlate with the subballast sample test results; some GPR interpretations indicated individual high substructure moisture contents in places where the substructure material was found to be dry and had a low fines content. This phenomenon was observed even when no GPR signal disturbing structures, such as frost insulation boards, were detected.
- Ballast sampling conducted in accordance with current Finnish guidelines did not give a true representation of the ballast material found in the structure, due to the fouled ballast material located lower than where the sample is to be taken.
- Future research should focus on depicting substructural conditions more explicitly in the initial data, especially the idea that parametrization of drainage conditions should be incorporated into future track maintenance data.

References

1. Li, D.: 25 years of heavy axle load railway subgrade research at the Facility for Accelerated Service Testing (FAST). *Transportation Geotechnics* 17, 51–60 (2018). doi:10.1016/j.trgeo.2018.09.003
2. Li, D., Hyslip, J., Sussmann, T., Chrismer, S.: *Railway geotechnics*. CRC Press, Taylor & Francis Group, Boca Raton (2016). doi:10.1201/b18982
3. Andrade, A., Teixeira, P.: A Bayesian model to assess rail track geometry degradation through its life-cycle. *Research in Transportation Economics* 36(2012) 1–8 (2018). doi:10.1016/j.retrec.2012.03.011
4. Andrade, A., Teixeira, P.: Statistical modelling of railway track geometry degradation using Hierarchical Bayesian models. *Reliability Engineering and System Safety* 142(2015) 169–183 (2015). doi:10.1016/j.res.2015.05.009
5. Vale, C., Lurdes, S.: Stochastic model for the geometrical rail track degradation process in the Portuguese railway Northern Line. *Reliability Engineering & System Safety* Vol. 116, s. 91–98 (2013). doi:10.1016/j.res.2013.02.010
6. Andrews, J., Prescott, D., Rozières, F.: A stochastic model for railway track asset management. *Reliability Engineering and System Safety* 130, 76–84 (2014). doi:10.1016/j.res.2014.04.021
7. Esveld, C.: *Modern railway track 2nd edn*. MRT-Prod, Zaltbommel (2001).
8. Shenton, M.: Ballast deformation and track deterioration. *Track technology*, 253–265 (1985).
9. Li, D., Selig, E.: Evaluation of railway subgrade problems. *Transportation research record*, 1489, 17–23 (1995).
10. Hájek, P., Havránek, T.: *Mechanizing Hypothesis Formation: Mathematical Foundations for a General Theory*. 1st ed. Springer, Berlin/Heidelberg (1978).
11. Rauch, J.: *Observational Calculi and Association Rules*. *Studies in Computational Intelligence* 469. Springer, Heidelberg (2013). doi:10.1007/978-3-642-11737-4
12. Berka, P.: Practical aspects of data mining using LISp-Miner. *Computing and Informatics* Vol. 35, 528–554 (2016).
13. Finnish Transport Infrastructure Agency (in Finnish: Liikennevirasto): *Ratatekniset ohjeet (RATO) osa 3 Radan rakenne, annex 5* (2018).
14. Arnold, G., Fon Sing, P., Saarenketo, T., Saarenpää, T.: Pavement moisture measurement to indicate risk to pavement life. *New Zealand Transport Agency research report 611*, p. 22 (2017).
15. Luomala, H., Rantala, T., Kolisoja, P., Mäkelä, E.: Assessment of track quality using continuous track stiffness measurements. In: *3rd International Symposium Railway Geotechnical Engineering: 23-24 November 2017*, pp. 218–289. IFSTTAR, Marne La Vallee, France (2017).
16. SFS-EN 933-1. Tests for geometrical properties of aggregates. Part 1: Determination of particle size distribution. Sieving method. (2012).
17. Finnish Transport Infrastructure Agency (in Finnish: Ratahallintokeskus): *Ratatekniset ohjeet ja määräykset (RAMO) osa 15 Radan kunnossapito* (2002).