American Journal of Operations Research, 2022, 12, 261-292
https://www.scirp.org/journal/ajor
ISSN Online: 2160-8849
ISSN Print: 2160-8830

Multivariate Analyses for Finding Significant Track Irregularities to Generate an Optimal Track Maintenance Schedule

Mami Matsumoto[1], Masashi Miwa[2], Tatsuo Oyama[3]

[1]Railway Technical Research Institute, Hikari-Cho, Kokubunji-Shi, Tokyo, Japan
[2]Tokyo Keizai University, Minami-Cho, Kokubunji-shi, Tokyo, Japan
[3]National Graduate Institute for Policy Studies, Roppongi, Minato-Ku, Tokyo, Japan
Email: matsumoto.mami.78@rtri.or.jp, miwa@tku.ac.jp, oyamat@grips.ac.jp

How to cite this paper: Matsumoto, M., Miwa, M. and Oyama, T. (2022) Multivariate Analyses for Finding Significant Track Irregularities to Generate an Optimal Track Maintenance Schedule. *American Journal of Operations Research*, **12**, 261-292.
https://doi.org/10.4236/ajor.2022.126015

Received: October 13, 2022
Accepted: November 26, 2022
Published: November 29, 2022

Abstract

We first discuss the relationship between the optimal track maintenance scheduling model and an efficient detection method for abnormal track irregularities given by the longitudinal level irregularity displacement (LLID). The results of applying the cluster analysis technique to the sampling data showed that maintenance operation is required for approximately 10% of the total lots, and these lots were further classified into three groups according to the degree of maintenance need. To analyze the background factors for detecting abnormal LLID lots, a principal component analysis was performed; the results showed that the first principal component represents LLIDs from the viewpoints of the rail structure, equipment, and operating conditions. Binomial and ordinal logit regression models (LRMs) were used to quantitatively investigate the determinants of abnormal LLIDs. Binomial LRM was used to characterize the abnormal LLIDs, whereas ordinal LRM was used to distinguish the degree of influence of factors that are considered to have a significant impact on LLIDs.

Keywords

Multivariate Analysis, Track Maintenance Scheduling, Track Irregularity, Longitudinal Level Irregularity Displacement, Cluster Analysis, Principal Component Analysis, Binomial Logit Regression Model, Ordinal Logit Regression Model

1. Introduction

The railroad business was privatized in Japan based on the decision of the Nakasone Cabinet in 1987. Currently, railway utilities in Japan are dominated by two major groups: the Japan Railway (JR), comprising six companies that are headquartered in Hokkaido, East Japan, Central Japan, West Japan, Shikoku, and Kyushu, and privately owned railways operated by general private utilities, comprising approximately 90 companies. The number of passengers transported by the JR and the private group was 6707 (10^6 passengers) and 10,963 (10^6 passengers), respectively, in fiscal year 2020, denoted by FY2020. The number of passenger-kilometers (sum of the products of the number of passengers and their traveled distance in kilometers) transported in FY2020 is 152,084 (10^3 passenger-km) and 111,127 (10^3 passenger-km) for the JR and private railways, respectively [1]. Passenger transportation volume of the railways has been following a downward trend since its peak in 2018. Recently, it has declined significantly due to the COVID-19 outbreak. These data in [1] indicate that the JR group is responsible for 38% of the total number of passengers traveled and 58% of the total number of passenger kilometers traveled in Japan. Based on these findings, it is true that railways can be expected to remain the most important and indispensable public transportation system in Japan in future, if we consider the size of the country and the current geographical population distribution. In addition, the railway system plays an important role in the world economy and the transportation of people. However, the business of railway utilities has suffered severely because of the declining birthrate, aging population, and decrease in the working population, in addition to the decrease in the number of passengers because of the COVID-19 outbreak; thus, a decline in railway business revenues can be seen in most railway utilities in Japan. In particular, we foresee that by 2040, the railway business revenues of the JR group will decrease by 20% compared with those in 2019 [2]. Therefore, improving the business condition by increasing efficiency is one of the important issues [3].

Preventing railway accidents and obtaining passenger safety are necessary to ensure transportation safety and increase the reliability of transportation networks. Compared with other modes of transportation, railways are superior in terms of reliability, running speed, and safety. In terms of safety, the probability of accidental death per passenger kilometer per hour is 1.5×10^{-11}, which is approximately 1/7 for airplanes and approximately 1/450 for automobiles. Thus automobiles have the highest probability of accidental death. The probability of accidental death per hour is approximately 1/2 for automobiles compared to that of airplanes, which have the highest probability of accidental death. However, the probability of death for railways is approximately 1/670, which is considerably lower than that of airplanes [4] [5].

However, some large-scale accidents such as the Shigaraki accident (42 dead, 614 injured) in 1991 and the Fukuchiyama accident (107 dead, 562 injured) in 2005 in which many passengers died, made a significant social impact, and the safety of the railway system became a major social issue. The social unrest and economic losses caused by these rare, large-scale accidents were extremely high. Railway utilities also suffered significant losses in terms of compensation for damages, loss of credibility, and reduction in transportation revenue. In addition, even small-scale accidents can impair railway operations and reduce transportation reliability. Thus, accidents prevent the railway system from fulfilling its role as a public transportation system.

In Japan, the railway system is a crucial public transportation system, and as a public utility, the railway business is among the most important businesses; hence, the safety of the railway system is critical. When considering the safety of railway systems, it is necessary to consider how track maintenance should be performed because track safety is the most essential concern. Hence, in this study, we proposed an efficient method for detecting longitudinal-level irregularity displacement (LLID) anomalies and validated the proposed method.

While playing an important role as social infrastructure, railway utilities are facing fierce competition from other transportation systems, and there is an urgent need for examining and verifying safety improvement strategies to ensure stable and safe transportation. It is necessary to select targets on which accident prevention measures should be focused, taking into account the impact of accidents and the scale of human, economic and social losses, and to examine ways to efficiently improve safety [6] [7]. Previous studies clarified the relationship between safety investment cost and effectiveness by optimizing the risk minimization and the life cycle cost minimization models [8] [9].

Many studies have been conducted to improve railway safety and prevent specific accidents. In particular, the study of large-scale accidents helps us develop and introduce new technologies and mechanisms to prevent the occurrence of similar accidents. There are two main approaches for accident prevention: one way is to identify an accident or event to be analyzed, estimate the process of accident occurrence, and study preventive measures against it. The other is to first divide the railway into several subsystems, extract as many hazards as possible, then estimate the risk posed by the hazards and eliminate those hazards that are judged to be high-risk hazards. The former research is based on micro-analysis, wherein the focus is on detailed accident causes and processes of accidents. The latter, on the other hand, is based on macro analysis, wherein the statistical data on the number of accidents, their causes, and the details of the accidents are first analyzed. Then, based on the analysis results, the overall trends are identified to find directions for developing and introducing new technologies and systems.

However, train accidents in Japan have not been investigated much, thus few studies have been conducted so far. [10] measured the safety of Japanese railway systems to mitigate train accidents by applying statistical methods, and proposed a model to explain the occurrence processes of train accidents from human viewpoints, such as train operators' judgments and behaviors. Furthermore, [11] proposed several countermeasures to reduce the number of serious train accidents (STAs). Using data from 1987 to 2005, [12] and [13] examined the STAs in Japan from various viewpoints such as their frequency distributions of occurrence, types, causes, and countermeasures. In particular, they showed statistical models and observed that the frequency and interoccurrence interval of STAs could be explained using the Poisson process. [14], using data from 2001 to 2011, investigated the occurrence of train accidents and their effects on train operations such as delays and cancellation of train services in Japan. [15]-[20] also discussed train accidents in Japan using historical data.

Train accident-causing factors in the United Kingdom have been investigated quantitatively in particular, where major accidents occurred frequently after the privatization of national railways in 1992. [21] and [22] investigated train accidents in Great Britain using 75 fatal train accident data points by classifying them based on accident types. He built a model to estimate the occurrence probability and the number of deaths due to train accidents; and concluded that introducing accident prevention equipment and devices is effective. These studies by Evans are one of the few examples of cross-sectional analyses of accident data classified by the accident type and factors leading to the accident. [23] developed a model considering the occurrence processes such as the birth process and Poisson processes to analyze the risk of derailment using a probabilistic approach. [24] demonstrated that the risk of derailment can be significantly reduced by increasing the inspection frequency of grinding related to rolling contact fatigue. [25] proposed a discrete-time Markov chain method for assessing the global risk of railway transportation systems. [26] described suitable techniques used to reduce train accidents and protect infrastructures by preventing railway suicides, trespassing, and level crossing (LX) users.

[27] described that safety was the core issue in railway operations, emphasizing that Railway LX safety was a highly critical aspect of railways, and performed a causal reasoning analysis of LX accidents using the Bayesian risk model. They provided results to improve countermeasures to reduce the risk and mitigate the consequences of LX accidents. [28] investigated various causing factors of railway accidents in Iran. They concluded that approximately 55% of those accidents are caused by derailment. [29], describing the problem of railway safety management and the criteria for evaluating traffic risk, built an assessment model of traffic risk at LX. Their model was based on the plan-do-check-act (PDCA) processes for Lithuanian railways. Assessing the safety of LX accidents by using the logistic regression method, they concluded that their risk evaluation model was flexible and could be easily adapted for evaluating and monitoring of the safety risks of other elements pertaining to railway infrastructure. [30] attempted to identify hazardous situations, causes of hazards, potential accidents, and severe resulting consequences.

In this paper, we propose an efficient method to detect abnormal LLID lots from an extraordinarily large data set. This method can help us generate an optimal track maintenance schedule as an efficient detection of abnormal LLID lots can reduce a large amount of data processing work and computation time for obtaining an optimal track maintenance scheduling solution. The rest of this paper consists of five sections. Section 2 discusses the relationship between the optimization model for an optimal track maintenance plan and an efficient detection method for an abnormal LLID. Additionally, the sample data collection and verification methods used in this analysis are described. In Section 3, we attempted to apply the cluster analysis technique to the sample data to efficiently detect the points to be monitored based on actual LLID data. In Section 4, we use the principal component analysis (PCA) technique to clarify the characteristics of the data contained in each cluster obtained in the previous step. In Section 5, binomial and ordinal logit regression models (LRMs) are used to quantitatively analyze the determinants of abnormal LLID and the degree of influence of factors. Finally, Section 6 provides a summary and conclusions of this study.

2. Optimal Track Maintenance Scheduling (OTMS) Model and Detection Method for Abnormal LLID Locations

2.1. Model Analyses for Generating an OTMS

Tracks are the basic infrastructure of railway systems; however, according to [31], the track maintenance cost accounts for almost one-third of the total railway operating costs. Moreover, because the number of passengers is likely to decrease as the population declines, the revenues of railway businesses are also expected to decline. In addition, the expected increase in labor costs due to the decrease in the labor force requires a fundamental review of the railroad business structure. To improve maintenance management performance under such circumstances, an effective and efficient track maintenance strategy must be determined and implemented.

Many studies have been conducted to generate optimal track maintenance schedules based on the representation of these processes using mathematical models. [32] and [33] provided comprehensive reviews of maintenance scheduling. In particular, [32] classified maintenance scheduling into three levels: strategic, tactical, and operational, according to track conditions and planning periods, which allowed maintenance management to be implemented at various levels. [34] proposed a proactive maintenance schedule to prevent unexpected breakdowns in railway systems. They presented a mathematical formulation by applying greedy heuristics and compared the performance of their heuristics with the optimal solution using randomly generated instances. [34] and [35] developed a scheduling model that considered both short-term maintenance and long-term projects to minimize the overall cost and optimize track maintenance to prevent railway accidents.

We know that optimizing maintenance actions at a preventive level can reduce the track maintenance costs during the track lifecycle. Thus, scheduling preventive maintenance actions is important for railway administrations. [17] [36] and [37] developed OTMS models that consider the risk of derailment accidents. In particular, [37] aimed to obtain an optimal maintenance schedule for improving railway track irregularities using all-integer linear programming (AILP) optimization model analyses. They predicted changes in surface irregularities by investigating the transition process through degradation and restoration model analyses. They finally obtained an optimal tamping schedule for the multiple tie tamper (MTT) for the entire year. [17] presented an OTMS model developed for maintaining adequate railway track conditions and accomplishing efficient management of the railway service. Their models have been utilized by several major Japanese railway companies to develop their own OTMS models. Based on these optimization models, systems for planning operational schedules for various maintenance machines ([38] [39]) or for combining different maintenance machines to perform more sustainable track maintenance ([40]) have been developed, constructed, and empirically investigated.

In track maintenance plan optimization, a large amount of data needs to be processed. Hence, the computational complexity required to obtain an exact optimal solution is enormous. Therefore, multifaceted efficiency procedures, such as methods for collecting a large amount of data, improving processing efficiency, and streamlining optimization calculations, have been proposed and verified. This paper deals with the issue of how to efficiently collect LLID data from the viewpoint of collecting a large amount of data and improving the data processing efficiency. A specific and efficient method for detecting abnormal LLID lots is also proposed.

[41] developed a mixed-integer linear programming (MILP) model to minimize the maintenance cost over the track life cycle while ensuring the required geometrical quality. Their model was formulated to optimize tamping operations in ballasted tracks for preventive maintenance. [42] presented an optimization model for scheduling the ballast, rail, and sleeper renewal operations at the network level. Their objective was to minimize the expected railway track life-cycle cost (LCC) and track unavailability costs, which were derived from user impacts caused by traffic disruption. In addition, [43] proposed a strategic model to optimize railway-track renewal operations at a network level aiming for a railway track geometry degradation that considers uncertainties in the forecast by defining a track geometry reliability parameter. Their model was formulated by applying a multi-objective optimization approach to assess railway track maintenance strategies taking a cost-reliability trade-off into consideration.

2.2. Detection Methods for Finding Abnormal LLID Locations

The final objective of this study is to develop a strategy to improve the safety of railway systems by optimizing track maintenance, which is an important issue in railway business management. Track maintenance management requires efficient determination of the location of an abnormal LLID. Therefore, track inspections are usually conducted periodically to perform planned maintenance and recover track geometry ([44] [45] and [46]). However, in rare cases, LLID can be localized and rapid owing to a large cavity in the roadbed, or initial settlement in the ballast after maintenance. This can threaten the safe operation of trains and may cause transport disorder and accidents. As a result, unplanned actions, such as immediate track maintenance, are performed. For stable train transportation, it is necessary to detect signs of rapid LLID growth, identify the location of its occurrence and its tendency at an early stage, and perform track maintenance in advance. However, detecting abnormal LLID lots quickly and efficiently has not been done much so far, and furthermore, detection method has not been investigated before. However, detecting abnormal LLID lots quickly and efficiently has not been attempted much so far, and furthermore, detection method has not been investigated before.

[47] proposed a computer application that uses track geometry measurements and maintenance operation data. It has been applied in France. [48] proposed a method to schedule track inspection and maintenance activities in railroad networks. [49] proposed a simplified condition-based maintenance (CBM) for implementing LLID and carrying out appropriate and timely maintenance on track locations. [50] proposed a prediction framework validated using LLID data obtained from measurements performed using track geometry cars. Their method showed a superior prediction performance than that of state-of-the-art methods. [51] developed a support vector machine (SVM) to model the deterioration of track geometry defects, focusing on three defect types: surface, cross-level, and dip. The model results demonstrate a prediction accuracy higher than 70%.

Recently, improvements in LLID measurement and analysis technology have enabled obtaining track inspection and acceleration data with high frequency and simplicity. [52] developed a prototype track-measuring device mounted on a commercial railway vehicle. [53] and [54] confirmed that for the diagnosis and prognosis of rail defects, rail corrugation can be managed efficiently using the leading axle of the bogie. [55] showed that the detection performance could be improved by combining multiple detectors within a multitask learning framework, and their approach resulted in improved accuracy for detecting defects. [56] reported that their proposed method could improve reliability and safety as they grasped the details of the LLID situation more clearly by visualizing the data. [57] proposed a decision-support approach for the CBM of rails based on expert-based systems. Their results supported infrastructure managers in analyzing the problems in their rail infrastructure and efficiently performing CBM decision-making. [58] utilized a track-recording vehicle for maintenance purposes, which has the potential to provide efficient and frequent measurements on a train car body by applying an augmented state Kalman filter. Some studies have examined the efficiency of maintenance, operation, and marketing using data envelopment analysis (DEA) in railway business management ([59] [60]).

[61] surveyed recent applications of machine learning (ML) for the diagnosis and prognosis of rail defects in rail track maintenance. They presented the shortcomings of current techniques and discussed what the research community and rail industry could do to address these issues and suggested future research directions. [62] developed a neural network model and revealed that the neural network technique is capable of establishing correlations between geometrical defects and structural problems in tracks. [63] introduced data-driven models, such as artificial neural networks and support vector regression, as basic ingre-

dients of ML technology. They concluded that the maintenance and renewal of ballasted tracks could be optimized in terms of time and cost by deriving an appropriate statistical model of track deterioration.

2.3. Sampling Methods for Data Collection

LLID is generally obtained by placing a reference string on the rail and measuring the distance between the center of a straight line connecting the two ends of the string and the rail at the center of the string. The LLID obtained when the length of the reference string is 10 m is called a 10 m-chord irregularity, which is widely used by Japanese railway utilities as an indicator for track maintenance. However, the 5 m-chord irregularity, which is obtained when the reference string length is 5 m, is known to detect anomalies more accurately than the 10 m-chord irregularity [64]. Therefore, the 5 m-chord LLID was mainly utilized in the analyses performed in this study. When developing the detection method for abnormal LLIDs, the following assumptions were made about the characteristics of the target line segment to be analyzed: 1) annual passing tonnage is approximately 1.7 to 2.5 million tons, 2) total track length is 558.9 km, 3) high-speed lines with no freight traffic (Shinkansen), and 4) track inspection period of frequency is once every 10 days. This data was collected from 2015 to 2020. In other words, the data were obtained for 168 days of inspection.

Generally, control indices are defined based on the 10 m-chord irregularity in order to maintain safety and a certain level of riding comfort, and to reduce the amount of emergent maintenance work. In this study, we defined three levels of control indices: the first is a maintenance target value for planning regular maintenance work; the second is an operative maintenance target value defined to maintain a certain level of riding comfort; and the third is an extreme displacement value; when this value is reached, maintenance work is immediately applied.

The total track length of 558.9 km was divided into lots of 100 m each, and the maximum value of LLID of the displacement, which is hereafter referred to as maximum displacement, was determined for each lot. Moreover, the set of 5589 lots is referred to as the total set of lots. The plotted distribution for the 5 m-chord and 10 m-chord cases of the maximum displacement is shown in **Figure 1**. As shown, the 10 m-chord maximum displacements tend to be larger than those of the 5 m-chord. However, we found that the difference is not significantly large because the coefficient of determination between the two datasets is 0.645.

In this analysis, 500 lots, which we call the sample set of lots, were selected out of the total set of 5589 lots. In sampling, we selected the sample set of lots such that their characteristics were as similar as possible to the total set. Accordingly, we considered the following eight categories related to track structure, rail characteristics, train operations, etc.: i) for rails, regular rail, and long rail with glued joints. ii) For the track bed, the concrete slab, ballast, direct connection to the sleeper, and boundary lot. iii) We classified the linear and curve radii into straight

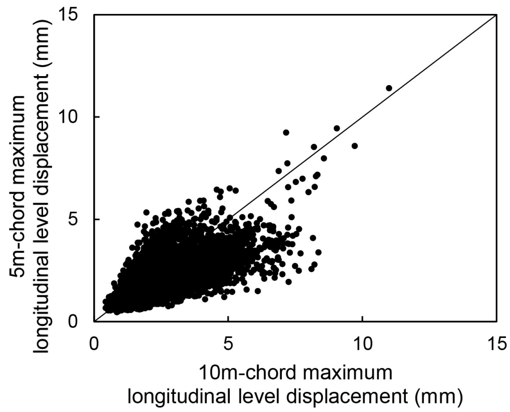


Figure 1. Distribution of the 5 m-chord and 10 m-chord cases of maximum displacements.

lines and curves. Curves were classified into three categories according to the curve radius, and composites and joints connected to these curves were added. iv) structure, v) running speed, vi) branching, vii) expansion joint, and viii) glued-insulated joints of rails were also selected so that the percentages of the total and the selected lots were as similar as possible.

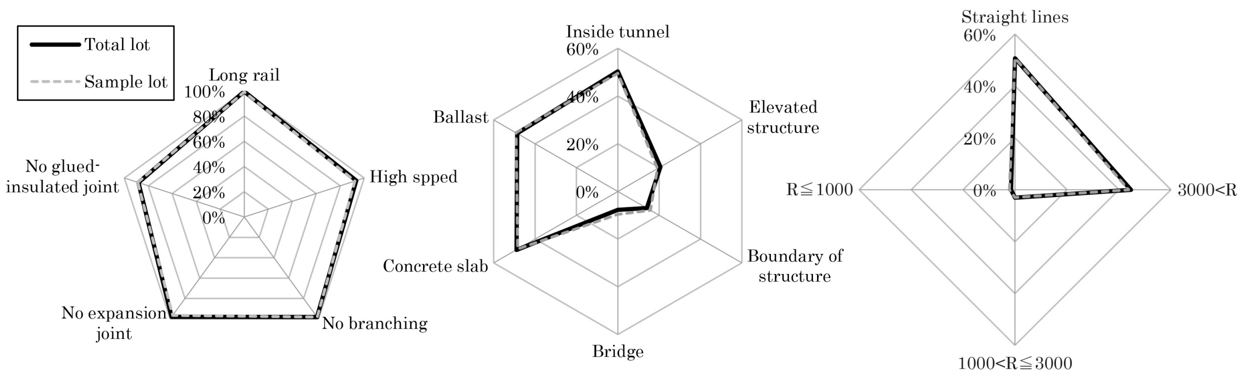
Regarding the rail, we considered categories ii), iii), and iv), whereas for the track structure, we distinguished the inside of the tunnel and the tunnel mine-head, and separated elevated structures, bridges, and the boundary areas of the structures. Train speeds v) were further categorized as high-speed and low-speed lots running at 170 km/h or higher and at less than 170 km/h, respectively. For branching vi), expressions 0, 1, and 2 were used, depending on the number of branches in the lot. Expansion joint vii) and glued-insulated joint viii) were set to 0 and 1, respectively, depending on whether they were located at the lot. It is well known that concrete slabs are widely used on high-speed lines because the occurrence of track irregularity displacement is unlikely, and the track conditions inside the tunnel tend to be stable because they are less affected by temperature changes than those at the mine-head of the tunnels. Moreover, the boundary lots of rails and structures, as well as expansion joints and glued-insulated joints, tend to be weak points compared to other general tracks; thus, track irregularity displacement with large maximum displacement is more likely to occur because they are difficult to maintain.

As mentioned earlier, we identify a sample set of lots such that its characteristics are similar to those of the total set. Figure 2 also shows the percentage distribution of the characteristics percentages of the total and sample sets of lots. It can be observed that the percentage distribution of both were almost identical.

3. Applying Cluster Analysis Technique to Classify Abnormal LLID Lots

3.1. Definition of Variables

Table 1 lists the variables used in this analysis. The variables were categorized



(a) Rail structure (b) Structure (c) Linear and curve radius

Figure 2. Percentage distribution for characteristics of the total and sample sets of lots.

Table 1. List of variables.

Category	Variables
I (Displacement-related)	MXDP, STDV, MXDV, NIMP, NRAD
II (Rail structure-related)	BRIN, BLST, NTBD, CRVR, EXJT, GIJT, RGRL
III (Operation-related)	RSPD, PTWT

into the following three types: In the following explanation bin. indicates binary variable,

(1) Track displacement-related: maximum irregularity displacement (MXDP, mm), standard deviation (STDV, mm), maximum deviation (MXDV, mm), number of improvements (NIMP), number of rapid advances (NRAD).

(2) Rail structure-related: branching (BRIN, bin.), ballast (BLST, bin.), concrete slab (NTBD, bin.), curve radius (CRVR), expansion joint (EXJT, bin.), glued-insulated joints (GIJT, bin.), regular rail type (RGRL, bin.).

(3) Train operation-related: running speed (RSPD, km/h), passing train weight (PTWT, ton/year).

For variables in (1), which are related to the track irregularity data, MXDP, STDV, and MXDV denote the maximum value, standard deviation, and maximum deviation of the measured LLIDs, respectively, for the 5m-chord in the 100 m lot. The NIMP indicates the number of times that the LLID improvement was 1 mm or more on consecutive days during the measurement period. Whereas, NRAD indicates the number of times that the LLID advance was 2 mm or more on consecutive measurement days. The rail structure-related characteristic variables in (2) are binary variables that are set to 1 when each characteristic is included and 0 otherwise. However, the curve radius was set to 1 when $R > 1500$ m and 0 when $R \leq 1500$ m. For the variables in (3), the running speed and passing tonnage were measured in km/h and ton/year, respectively.

For each of the MXDP, STDV, and MXDV variables, we know that the larger the values, the worse could be the track condition. The NIMP indicates the number of maintenance cycles, whereas MXDV and NRAD are considered to indicate the tendency of the track condition to deteriorate rapidly (or to improve significantly after a single maintenance cycle). In general, the worse the track condition, the higher is the maintenance frequency. Therefore, there was a high correlation between MXDP, STDV, and MXDV. The NIMP was defined to be consistent with the maintenance frequency calculated from the maintenance results. The definition of the NRAD was based on the fact that the difference between the operative maintenance target value and the extreme displacement value among the control indices was 2 mm. Therefore, these indicators effectively represent the line segments characteristics considered in this analysis, and it is necessary to consider appropriate values when applying them to line segments under different conditions.

3.2. Numerical Results of Applying Cluster Analysis Technique

We present the numerical results of applying cluster analysis to a set of 500 lots in the sample set. The five variables used here are included in Category I, *i.e.*, MXDP, STDV, MXDV, NIMP, and NRAD. The data were classified into five clusters based on the *k*-means method. In applying the cluster analysis techniques, the variables, units, and notations listed in **Table 1** were used.

Table 2 shows the number of lots in the sample set included in each cluster and the center-of-gravity coordinates, that is, the average of all corresponding values for the variables contained in each cluster. Clusters {1, 2, …, 5} are denoted as {C1, C2, C3, C4, C5}. In **Table 2**, the center-of-gravity coordinates of C1 are the largest for almost all variables given by {MXDP, STDV, MXDV, NIMP, and NRAD}. We can see that the variable values for C1, C2, C3, C4, and C5 decrease in the same order, except in the case of variable NIMP, where the magnitudes of the values for clusters 2 and 3 are reversed. This suggests that the magnitude of LLID is in the order of C1, C2, C3, C4, and C5, and that the sample set of lots in C4 and C5 represents the majority of the total set, approximately 90%, and has the smallest amount of LLID. Furthermore, the results in **Table 2** show that almost 70% of the 500 sample sets of lots were in C5, where the center-of-gravity coordinates had the lowest LLID values.

Table 3 lists the distances of the center of gravity coordinates among the five

Table 2. Number of lots by cluster and center-of-gravity coordinates by variables.

Cluster	NELM	MXDP	STDV	MXDV	NIMP	NRAV
1	6	11.370	1.837	9.607	11	2
2	27	9.436	1.746	7.696	4	0
3	19	8.112	1.374	6.024	8	0
4	116	5.843	1.026	4.070	2	0
5	332	2.481	0.206	0.957	0	0
Total	500					

NELM: Number of elements contained in each cluster.

Table 3. Distances between centers-of-gravities for clusters.

Cluster \ Cluster	1	2	3	4	5
1	0	7.168	5.598	11.739	16.381
2	-	0	4.685	5.598	10.591
3	-	-	0	6.951	11.192
4	-	-	-	0	5.025
5	-	-	-	-	0

clusters. From this result, it can be seen that the Cluster 5 is farther away from C4, C3, C2, and C1, in the given order. Moreover, it also shows that among these five clusters, C1, C2, and C3 are quite close to each other, whereas C4 and C5 are much farther from these and closer to each other. Thus, we can also see that they can be divided into two groups, {C1, C2, C3} and {C4, C5}. It should be noted that we argue that the lots in the first group {C1, C2, C3} obtained are those that should be targeted for detection as abnormal LLID (or MXDP) lots. We know that the shorter (longer) the distance from the center-of-gravity coordinates of the cluster, the higher (lower) is its similarity to the center. Thus, we find that clusters {C1, C2, C3} contain lots with large fluctuations and high maintenance frequencies, whereas {C4, C5}, which contain majority of lots, have lots with stable small track irregularities. Hence, these lots have insignificant improvement frequencies.

We aim to determine the track structure characteristics of each lot in the cluster. Table 4 shows the percentage of lots in each cluster based on track structure characteristics. The characteristics found in the lots contained in each of the clusters {C1, C2, C3} in the first group defined above are as follows.

All lots in C1 recorded large values of approximately 10 mm for MXDP, which correspond to significant values for the maximum value of the 10m-chord during the inspection period. Moreover, all lots are ballasted, have a large fluctuation in MXDP, and have several NIMP cycles. In addition, half of the lots are located at the boundaries of the structures, and many of them include EXJT or GIJT, which tend to deteriorate the track condition. Therefore, C1 is the group that requires the most attention for maintenance operations because it includes lots that are maintained at a high frequency and have a large NIMP and NRAD values. In addition, all lots are located in sections with curve radii of 3000 m or more, or in straight sections with high-speed lots. This suggests that the high-speed lots in this line are due to the fact that the track condition is more likely to deteriorate. Moreover, C2 contains lots with MXDP values close to extreme values. Compared to C1, C2 tended to have more lots with a lower maintenance frequency, even though the track condition was poor. This is because several lots in C2 are difficult to maintain on a routine schedule and require large-scale maintenance work such as track structures, structure boundary areas, and bridges.

Table 4. Percentages of track structures for each cluster (%).

Track	Item	C1	C2	C3	C4	C5
Rail	Long rail	100.0	96.3	100.0	97.4	100.0
	Unit length rail	0.0	3.7	0.0	2.6	0.0
Roadbed	Concrete slab	0.0	0.0	5.3	16.4	66.9
	Ballast	100.0	81.5	84.2	81.0	32.5
	Boundary	0.0	18.5	10.5	2.6	0.6
Liner and curve radius (R)	Straight lines	83.3	70.4	52.6	50.9	48.5
	$R > 3000$	16.7	29.6	47.4	41.4	46.1
	$1000 < R \leq 3000$	0.0	0.0	0.0	3.4	3.3
	$R \leq 1000$	0.0	0.0	0.0	4.3	1.8
Structure	No structure	16.7	3.7	15.8	10.3	1.2
	Inside tunnel	16.7	7.4	10.5	21.6	65.7
	Mine-head	0.0	0.0	5.3	4.3	0.6
	Elevated structure	0.0	29.6	21.1	26.7	16.0
	Bridge	16.7	33.3	26.3	16.4	4.2
	Boundary	50.0	25.9	21.1	20.7	12.3
Speed	High speed	100.0	96.3	100.0	89.7	92.2
	Low speed	0.0	3.7	0.0	10.3	7.8
Branching	No branching	100.0	96.3	89.5	96.6	99.4
	With 1 branching	0.0	3.7	10.5	3.4	0.6
Expansion joint	No expansion joint	83.3	88.9	84.2	92.2	99.7
	Included	16.7	11.1	15.8	7.8	0.3
Glued-insulated joint	No glued-insulated joint	66.7	85.2	78.9	89.7	89.8
	Included	33.3	14.8	21.1	10.3	10.2

Such lots often require some preparation period before maintenance can be performed. Lots in C2 also belong to a group that requires attention to rapid advances. Lots in C3 also occasionally have large MXDP with a large NIMP. Although several lots include BRIN, EXJT, and GIJT, which are difficult to maintain, they are maintained comparatively regularly. This may be due to the fact that most lots are managed on site as they are easily affected by the MXDP.

Regarding the characteristics found in the lots in each of the clusters {C4, C5}, C4 mainly contains lots with relatively stable track conditions, although there are some lots that have some MXDP fluctuations and several NIMP. This is due to the fact that many of the lots have sharper curves than the other clusters, which results in a smaller irregularity displacement advance due to slower train running speeds and less impact to the track by the trains. C5, which contains more than half of the sample set of lots, can be said to show the least fluctuation in MXDP among all clusters. Most lots had no maintenance record during the measurement period, as the MXDP remained close to 0 mm without fluctuation. More than half of the lots are located on the concrete slab, where irregularity displacement does not occur, or in tunnels, which are less affected by temperature changes. Therefore, this group had a relatively steady track condition.

3.3. Efficient Method for Detecting Abnormal LLID

Using the historical track inspection data of the LLID, five variables: MXDP, STDV, MXDV, NIMP, and NRAD were calculated for each 100 m lot during the measurement period, and a cluster analysis technique was applied to categorize the lots and characterize each cluster element. Based on the categorization, it was possible to divide the five clusters into two groups: one requiring attention to track management and the other that did not. The lots in the first group {C1, C2, C3} are considered to be those where abnormal LLID lots are likely to occur. Based on these analysis results, we describe a method for efficiently detecting abnormal LLID lots.

First, when detecting abnormal LLID lots, we attempted to find a lot in the sample set for which a large irregularity displacement or its advance actually occurred. Here again, the control indices based on the value of the 10 m-chord LLIDs are applied. In selecting the abnormal LLID lots, the following three indicators were used: (i) lot exceeding 10 mm, (ii) lot with 4 mm/10 days advance, and (iii) lot with the extreme displacement value. Indicator (i) is a lot in which the 10 m-chord MXDP exceeds the extreme displacement value of 10 mm. The operative maintenance target value is 6 mm and the extreme displacement value is 10 mm in the control indices for the 10 m-chord LLID. Therefore, Indicator (ii) is a lot whose MXDP has reached the operative maintenance target value of 6 mm, and its maximum value reaches the extreme displacement value by the next inspection after 10 days. Indicator (iii) is a lot for which the MXDP of 11.2 mm is actually detected on the 10 m-chord irregularity displacement, and the extreme displacement value is exceeded owing to rapid advancement.

As for the abnormal LLID lots, among the lots in the sample set, nine lots belong to Indicator (i), and one lot each belongs to Indicators (ii) and (iii), thus amounting to a total of 11 lots. The distribution of the 11 lots with abnormal LLID lots was five lots in C1 and six lots in C2. Thus, it can be seen that all abnormal LLID lots are in either C1 or C2. The lots in C1 tend to have a consistently large LLID among the abnormal LLID lots and need to be frequently improved, which implies that these lots have been periodically maintained. In C1, five of the six lots were abnormal LLID lots, but the remaining lot also detected MXDP of 9.6 mm and a large NIMP, thus showing a behavior similar to that of Indicator (i). On the other hand, C2 includes lots with steady periods of irregularity displacement progression but tends to occasionally suffer from rapid advances. As can be seen from the cluster analysis results, the distance between the center-of-gravity coordinates of C2 and C3 was close, and both the displacement data characteristics and structural characteristics were found to be similar. Furthermore, the lots in C3 also have a large number of improvements, and large displacements occur occasionally. Moreover, C3 has more lots that need maintenance more frequently. This tends to include lots that are maintained relatively regularly and managed in the field as locations where large LLIDs are likely to occur.

Based on these verification results, we conclude that lots included in C1 or C2, where a large LLID or rapid advance has actually occurred, should be carefully maintained. Furthermore, even if the MXDP is not extremely high, such as in C3, we should pay attention to lots with large NIMP and those including BRINs, EXJTs, and GIJTs. We consider that it is effective to prevent the occurrence of extreme values and post-maintenance by paying particular attention to the changes in the LLID and material conditions around the lots included in C1.

4. Applying Principal Component Analysis (PCA) for Finding Factors Related to the Occurrence of Abnormal LLID Lots

4.1. PCA for Applying the Sampling Data

PCA is a widely used method for extracting major features from big data. The main idea of PCA is to transform p variables of interest (x_1, x_2, ... x_p) into the same (or a smaller) number of variables (y_1, y_2, ... y_p), in descending order of their "importance". Thus, we express y_1, referred to as principal component (PC) 1, and denoted by PC1 as a linear combination. It is expressed as follows:

$$y_1 = a_{11}x_1 + a_{21}x_2 + \cdots + a_{p1}x_p, \tag{1}$$

where the coefficients a_{11}, a_{21}, ... a_{p1} are determined to maximize the variance of y_1 in (1) under the following condition:

$$a_{11}^2 + a_{21}^2 + \ldots + a_{p1}^2 = 1 \tag{2}$$

within the $p - 1$ dimensional space orthogonal to y_1, PC2 (y_2) and the coefficients a_{12}, a_{22}, ..., a_{p2} satisfying the same condition as Equation (2) can be determined to maximize the variance. Mathematically, PCAs can be calculated as the eigenvectors of the covariance matrix. The "importance" of each component, referred to as the proportion of variance, can be calculated as the corresponding eigenvalue divided by the sum of all eigenvalues. By applying PCA, we can find the principal features behind the entire dataset regarding how those data are distributed to each other.

Our methodology, which applies cluster analysis to the variables related to displacement, rail structure, and operation, which are given in **Table 1**, identifies regional characteristics and their principal features. This helps railway utilities to detect abnormal LLID lots and to more quickly apply their track maintenance operations to improve their operational performance. We consider that by taking the regional characteristics results into consideration, the PCA approach clarifies the principal factors behind the data of those variables, thus improving those factors could significantly reduce the LLID.

To detect abnormal LLID lots, PCA is applied using data based on the 14 variables shown in **Table 1** to identify background factors related to its occurrence and rapid advance process. The analysis was performed in two cases: using data from all 500 lots (Case 1) and using only 52 lots from C1, C2, and C3, which were determined to have bad track conditions by cluster analysis (Case 2). Note that the 52 lots in Case 2 were all linear sections, so there were 13 variables, excluding the "curve" variable CRVR. By calculating the correlation coefficients (CCs) among the 14 variables using data from all 500 lots, we found that the variables MXDP, MXDV, and STDV have very high CC values that are mutually higher than 0.91. The variables NIMP and NRAD also have rather high CC values, and are higher than 0.75 and 0.33, respectively. The other variables had low or small negative CC values. In general, when the raw data represent ratios or rates, if no large disparity exists between them, normalization of the data will probably expand the disparity. The eigenvalues of the CC matrix correspond to the standard deviations of the PC scores and represent the amount of information in the PC.

The PC load factor (L.F.) was obtained from the coefficient of the eigenvectors of each variable multiplied by the standard deviation of the eigenvalues for the four PCs. Here, the PC L.F. is obtained by multiplying the eigenvector coefficient by the square root of the eigenvalue.

4.2. Numerical Results of the PCA

The numerical results for Case 1 of the eigenvalues (E.V.s), where the PCA was performed on 14 variables for the sample set with 500 lots, are shown in **Figure 3** as a scree plot. The contribution rates (C.R.s) of E. V. s, and their cumulative contribution rates (C.C.R.s) of the CC matrix among the 14 variables are listed in **Table 5**. The E.V.s after rotation, based on the varimax method, and their

Table 5. E.V.s, C.R.s and C.C.R.s (Case 1).

P.C.	E.V.	C.R.	C.C.R.
1	4.605	32.892	32.892
2	2.157	15.405	48.297
3	1.613	11.524	59.821
4	1.278	9.128	68.949
5	0.888	6.345	75.294
6	0.834	5.957	81.251
7	0.756	5.402	86.653
8	0.641	4.581	91.234
9	0.477	3.404	94.639
10	0.373	2.662	97.301
11	0.247	1.767	99.067
12	0.068	0.485	99.552
13	0.043	0.305	99.857
14	0.020	0.143	100.000

E.V.: eigenvalue; C.R.: contribution rate; C.C.R.: cumulative contribution rate.

C.R.s are listed in Table 6. In this analysis, the E.V.s were approximately 1.0 or higher, and the C.C.R.s corresponding to the C.R.s were approximately 68.95% when we included the four PCs. Note that the C.R. is obtained by the sum of the squared load factor. Scree plots of the 14 E.V.s, and the four E.V.s after rotation (for those with values greater than 1) are shown in Figure 3. The L.F.s by variable for each of the four PCs are presented in Table 7. In this analysis, the E.V.s were approximately 1.0 or higher, and the C.C.R.s corresponding to the C.R.s were approximately 68.95% when we included four PCs. A graphical representation of Table 7 in descending order for each PC is also shown in Figure 4.

Next, the numerical results for Case 2 of the E.V.s, where the PCA was performed on 13 variables of 52 lots, are shown in Figure 5 as a scree plot. The E.V.s, C.R.s and C.C.R.s of the correlation matrix among the 13 variables are listed in Table 8. The E.V.s after rotation, based on the varimax method, and their contributions are listed in Table 9. In this analysis, the E.V.s were approximately 1.0 or higher, and the C.C.R.s corresponding to the C.R.s were approximately 74.67% when we included five PCs.

Table 6. E.V.s, C.R.s and C.C.R.s (Case 1).

P.C.	E.V.	C.R.	C.C.R.
1	4.128	29.48	29.48
2	2.165	15.46	44.94
3	1.776	12.69	57.63
4	1.585	11.32	68.95

Table 7. L.F.s by variable for each of the P.C., E.V.s, C.R.s and C.C.R.s (Case 1).

Variables	P.C.			
	1	2	3	4
MXDP	0.921	0.112	−0.011	0.190
STDV	0.898	0.229	−0.067	0.114
MXDV	0.936	0.196	−0.057	0.124
NIMP	0.879	0.090	−0.026	0.108
NRAV	0.597	−0.166	0.082	−0.187
BRIN	0.026	−0.024	0.074	0.764
BLST	0.415	0.832	0.091	0.133
NTBD	−0.348	−0.855	−0.092	−0.140
CRVR	−0.031	0.128	0.809	−0.019
EXJT	0.177	0.046	0.021	0.591
GIJT	0.232	−0.109	0.471	−0.395
RGRL	0.067	−0.043	0.493	0.503
RSPD	0.180	−0.129	−0.785	−0.254
PTWT	−0.222	0.743	0.046	−0.171

Table 8. E.V.s, C.R.s and C.C.R.s (Case 2).

P.C.	E.V.	C.R.	C.C.R.
1	3.032	23.33	23.33
2	2.410	18.54	41.86
3	1.752	13.47	55.34
4	1.417	10.90	66.23
5	1.097	8.44	74.67
6	0.885	6.81	81.48
7	0.667	5.13	86.62
8	0.551	4.24	90.86
9	0.480	3.69	94.55
10	0.435	3.34	97.90
11	0.205	1.58	99.47
12	0.043	0.33	99.80
13	0.026	0.20	100.00

Table 9. E.V.s, C.R.s and C.C.R.s (Case 2).

P.C.	E.V.	C.R.	C.C.R.
1	2.711	20.85	20.85
2	2.631	20.24	41.09
3	1.695	13.04	54.13
4	1.452	11.17	65.30
5	1.219	9.37	74.67

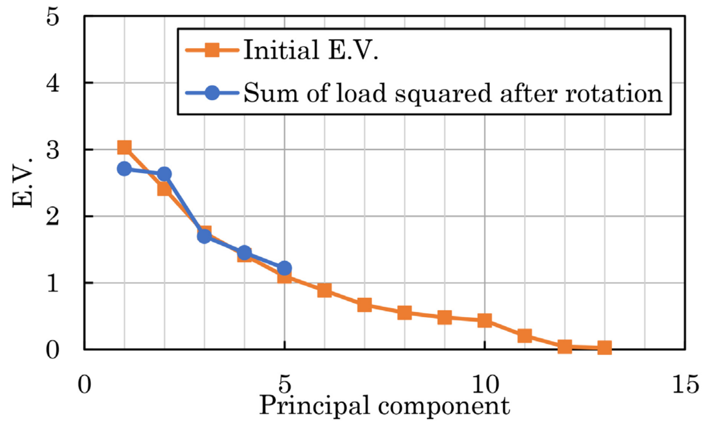


Figure 3. Scree plot of E.V.s (Case 1).

Scree plots of the 13 E.V.s, and the five E.V.s after rotation (for those with values greater than 1) are shown in **Figure 5**. **Table 9** lists the E.V.s, C.R.s, and C.C.R.s for Case 2. The L.F.s by variable for each of the five PCs, which were obtained from the eigenvectors or coefficients of the eigenvectors, are shown in

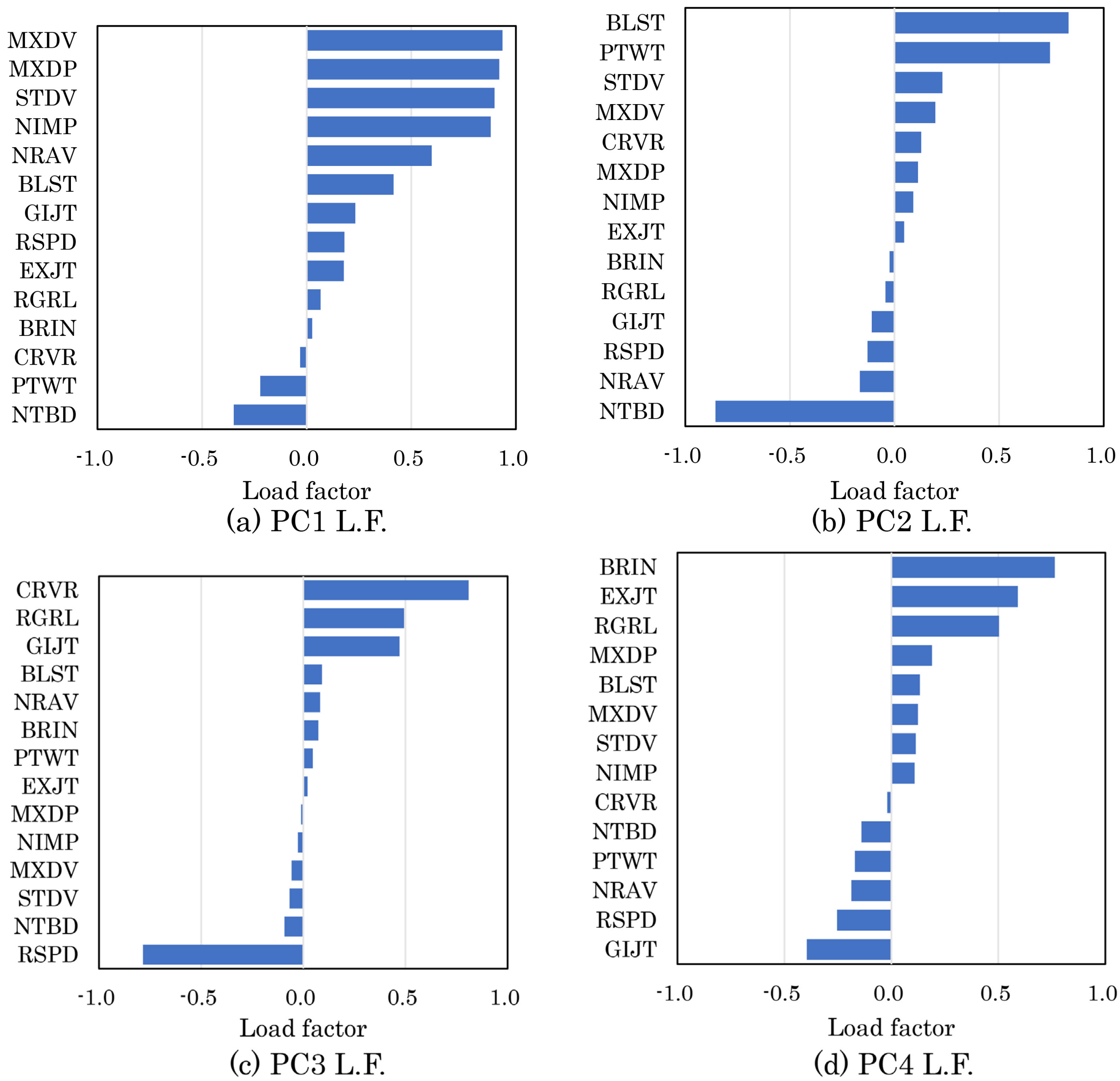


Figure 4. L.F.s by PCs (Case 1).

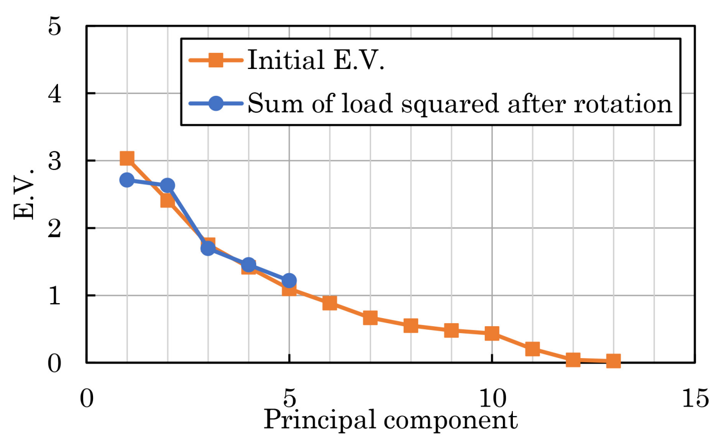


Figure 5. Scree plot of E.V.s (Case 2).

Table 10. A graphical representation of Table 10 in descending order for each PC is shown in Figure 6.

From Figures 4(a)-(d) and Figures 6(a)-(d), representing the descending order of PC L.F.s obtained from the coefficients of eigenvectors in Table 7 and Table 10, the following suggestions can be obtained for each PC. For PC1 in Case 1 and Case 2, we find from Figure 4(a) and Figure 6(a) that the L.F.s show that the LLID-related variables MXDV, MXDP, and STDV have large positive values, whereas those related to rail structure and operation: NTBD and PTWT, have small or negative values in Case 1. Moreover, those related to rail equipment, maintenance, and operation: EXJT, NIMP, and PTWT, have small or negative values in Case 2. Thus, we can conclude that PC1 in both Cases 1 and 2 represents the axis from the LLID data-related values to those related to rail structure, equipment, and operation.

For PC2 in Case 1 and Case 2, as shown in Figure 4(b) and Figure 6(b), the L.F.s show that variables related to rail structure, equipment, and operation: BLST, PTWT, RGRL, and BRIN, have large positive values, whereas the other variables related to rail structure and operation: NTBD and RSPD, have large negative absolute values. Thus, we can conclude that PC2 in both Cases 1 and 2 represents the axis from the values related to rail structure and operation to those related to maintenance and operation.

For PC3 in Case 1, as shown in Figure 4(c), the L.F.s show that variables related to rail structure, and the equipment: CRVR, RGRL, and GIJT, have large positive values, whereas the operation related variable RSPD has large negative absolute values. Thus, we can conclude that PC3 in Case 1 indicates the axis from

Table 10. L.F.s by variable for each of the P.C. (Case 2).

Variables	Principal Component				
	1	2	3	4	5
MXDP	0.864	0.075	0.197	−0.081	0.069
STDV	0.828	−0.196	−0.156	0.155	−0.203
MXDV	0.954	−0.021	0.033	0.077	0.108
NIMP	−0.178	−0.108	0.831	0.085	−0.018
NRAV	0.267	−0.019	0.778	−0.216	−0.053
BRIN	−0.182	0.743	−0.027	0.104	−0.210
BLST	0.158	−0.026	−0.277	0.760	−0.019
NTBD	0.028	−0.094	−0.147	−0.808	−0.212
EXJT	−0.341	0.543	0.161	0.286	−0.245
GIJT	0.234	−0.082	0.370	0.027	0.523
RGRL	0.094	0.934	−0.123	−0.038	0.071
RSPD	−0.071	−0.915	0.091	0.068	−0.221
PTWT	−0.148	0.048	−0.220	0.179	0.825

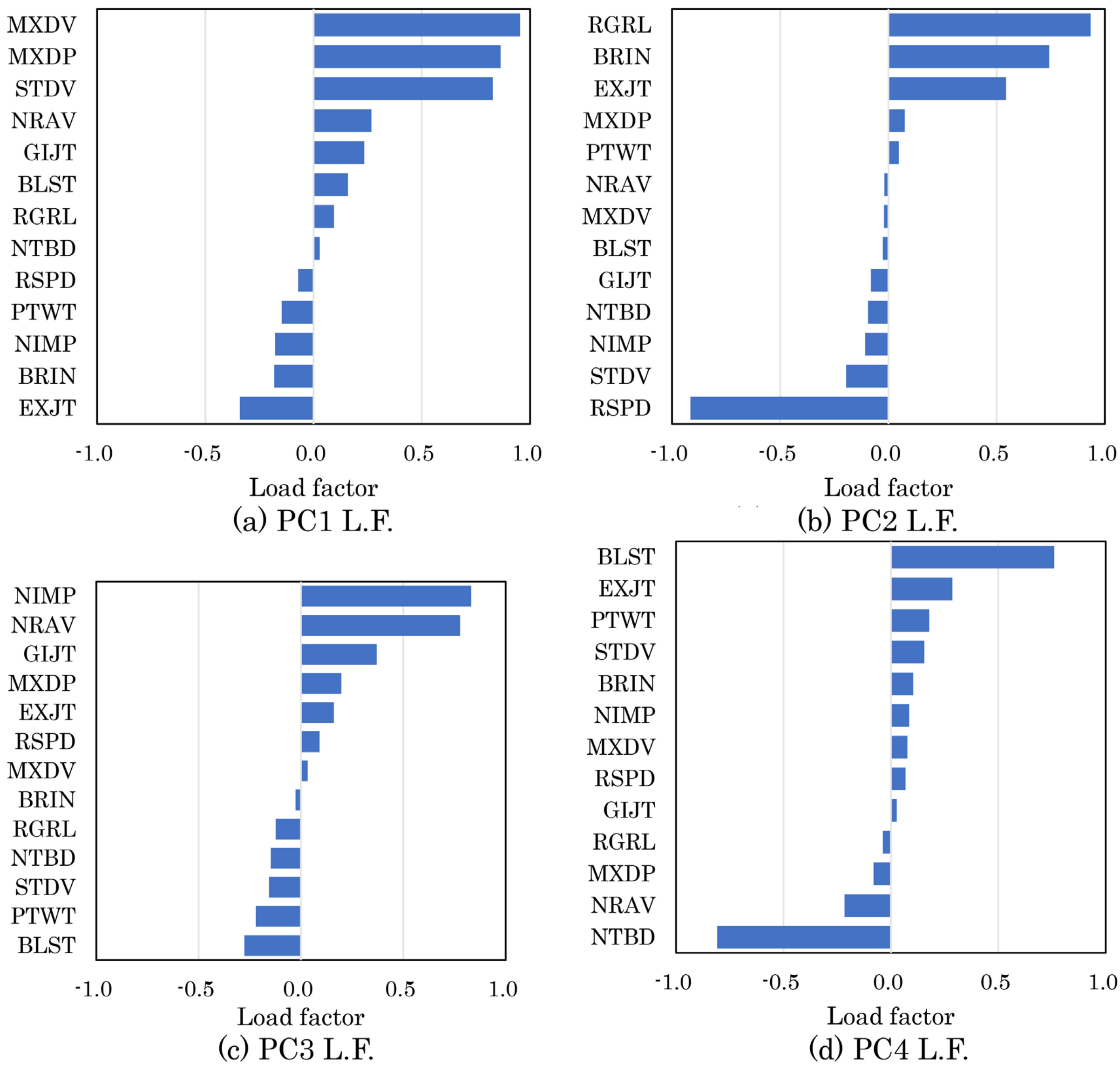


Figure 6. L.F.s by PCs (Case 2).

values related to rail structure and equipment to operation-related values. In Case 2, we find from **Figure 6(c)** that the load factors show that the displacement-related variables NIMP and NRAD have large positive values, whereas the those related to rail structure and operation: BLST and PTWT have large negative absolute values. Thus, we can conclude that PC3 in Case 2 represents the axis from maintenance-related values to those related to rail structure and operation.

For PC4 in Case 1, as shown in **Figure 4(c)**, the L.F.s indicate that rail equipment-related variables BRIN, EXJT, and RGRL have large positive values, whereas the other variables related to rail equipment and operation: GIJT and RSPD, have large negative absolute values. Thus, we can conclude that PC4 in Case 1 indicates the axis from values related to rail structure and equipment to operation-related values. In Case 2, we find from **Figure 6(c)** that the load factors show that the variables related to rail structure and equipment: BLST and EXJT, have large positive values, whereas the other variables related to rail structure and maintenance: NTBD and NRAD, have large negative absolute values. Thus, we can conclude that PC4 in Case 2 indicates the axis from the values related to rail structure and equipment to the other values related to rail equipment and maintenance.

Figure 7 shows a scatter plot of the scores for the principal components PC1 and PC2 for all variables. The orange points in Figure 7 indicate the results of Case 1 for all datasets from C1 to C5, whereas the blue ones correspond to those variables in Case 2, which only includes C1, C2, and C3. The cumulative contribution ratios of these PCs were 44.94% and 41.09%, respectively. We find from Figure 7 that those L.F.s are very close to positive values, whereas they are relatively farther from negative values. This is due to the difference in interpretation for PC1 and PC2, that is, variables with large positive and negative absolute values are similar for PC1, whereas they are dissimilar for PC2.

5. Applying Logit-Type Regression Model (LRM) Analyses to Investigate the Occurrence Mechanism of the Abnormal LLIDs

As described in Section 3.2, we classified the sample set of lots into five clusters with two major groups by applying a cluster analysis technique to 500 sample sets of lots: the set of lots with abnormal LLID in {C1, C2, C3} consisting of 52 lots and the set of other "normal" lots in {C4, C5} consisting of 448 lots. In this section, a quantitative analysis of the factors that cause abnormal LLID and actions required to reduce them are analyzed by focusing on the factors that cause abnormal LLID lots and by applying various LRMs. The lots that should be subjected to track maintenance in {C1, C2, C3} are further characterized in detail with respect to their influencing factors. Specifically, the set of lots to be subject to track maintenance obtained as a result of the cluster analysis was divided into

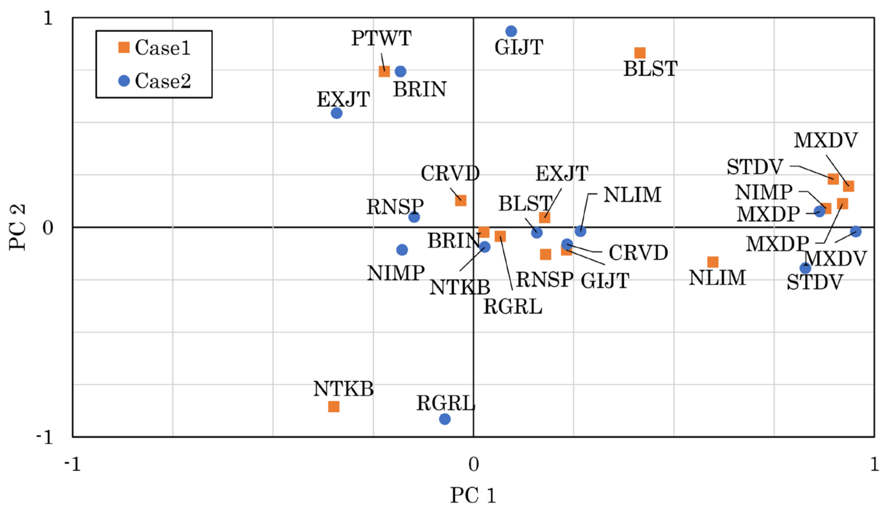


Figure 7. Scatter plot of P.C. scores PC1 and PC2.

three clusters with sets of 5, 27, and 19 lots in each cluster, and a quantitative analysis was conducted to determine the relationships among the factors affecting the occurrence and expansion of track irregularity. The purpose of this study was to apply ordinal LRM analysis.

5.1. Applying Binomial LRM Analysis for Identifying the Factors Causing Abnormal LLID

We developed an LRM to predict the occurrence probability of certain events. We assume that the log-odds function can be expressed by a linear regression model as follows:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n , \tag{3}$$

where x_1, x_2, …, x_n are independent variables, y is a 0 - 1 type dependent variable, a_0 (constant), and a_1, a_2, ..., a_n are parameters.

First, we developed a prediction model to estimate the probability that a certain event occurs by applying a logistic regression model. The logistic model corresponding to (3) is defined as follows:

$$f(p) = \ln\left(\frac{p}{1-p}\right), \tag{4}$$

where p represents the probability that a certain event occurs and $f(p)$ is the log-odds function, as the ratio $\frac{p}{1-p}$ denotes the odds indicating the ratio of the probability of the event occurring to that of the event not occurring. Then, assuming that the log-odds function (4) can be expressed by the linear regression model given in (3), that is,

$$f(p) = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n , \tag{5}$$

we obtain the following relationship from (4) and (5):

$$\ln\left(\frac{p}{1-p}\right) = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n . \tag{6}$$

The log-odds function given by (6) can be used to estimate the probability of occurrence of a certain event. Thus, the probability p can be expressed as

$$p = \frac{1}{1+\exp(-z)}, \tag{7}$$

where we assume that

$$z = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n . \tag{8}$$

All independent and dependent variables were categorical numbers rather than cardinal ones. The dependent variables are 1 and 0 depending on whether they belong to a certain category.

We developed a binomial LRM to quantitatively determine the influence factors related to the lots that should be subject to track maintenance, with the aim of detecting abnormal LLID that increases track irregularity as early as possible.

Variable selection was done using the PCA method and was based on the results of an analysis of the 52 lots in Case 1 ({C1, C2, C3}) in terms of track structure characteristics and operation characteristics. In the binomial LRM analysis, the dependent variables were 1 and 0 for lots in Cases 1 and 2, respectively ({C4, C5}). As independent variables, we used the 10 variables: MXDP, STDV, MXDV, NIMP, NRAD, BLST, NTRB, CRVR, RSPD, and PTWT, which were identified in the PCA as factors significantly affecting the occurrence and rapid advancement of track irregularity displacement. Table 11 presents the numerical results of the binomial LRM based on the dependent and independent variables described above. Using the 11 variables case as the initial model, Table 12 was obtained by reducing the variables so that the significance probability of each variable was less than 0.1. Table 12 shows the regression coefficient estimates, standard errors, Wald's square, significance probability, and odds ratio values. The Wald statistic value is given by the formula (estimate of partial regression coefficient/standard deviation of partial regression coefficient)^2 and follows a χ^2 distribution with one degree of freedom. Therefore, the larger the Wald statistic value, the higher is the dominance, and the greater is the influence on the dependent

Table 11. Parameter estimation results for the binomial LRM (11 variables).

Variable	z	Standard error	Wald square	Significance probability
MXDP	1.312	2.486	0.279	0.598
STDV	−4.557	6.456	0.498	0.480
MXDV	5.915	3.923	2.274	0.132
NIMP	3.350	1.919	3.047	0.081
NRAD	1.524	16.821	0.008	0.928
BLST	19.72	1229.5	0.000	0.987
NTRB	9.753	196.91	0.002	0.960
CRVR	−10.93	8151.63	0.000	0.999
RSPD	0.038	0.047	0.658	0.417
PTWT	0.000	0.000	0.134	0.714
CNST	−82.40	1230.50	0.004	0.947

Table 12. Parameter estimation results for the binomial LRM (5 variables).

Variable	z	Standard error	Wald square	Significance probability	Exp(B)	Odds Ratio
MXDP	2.924	2.472	1.400	0.237	18.62	18.62
MXDV	5.366	2.948	3.313	0.069	214.03	214.03
NIMP	4.223	2.229	3.590	0.058	68.21	68.21
RSPD	0.050	0.053	0.905	0.341	1.05	1.05
CNST	-81.506	48.503	2.824	0.093	0.00	0.00

variable. On the other hand, (7) shows that the smaller the significance probability, the larger the impact on the dependent variables. We add that the significant probability can be obtained by (7) using exp(z) values in **Table 11**. The odds ratio is based on the probability value p given by Equations (6), (7), and (8), and is expressed as

$$\text{odds ratio} = \frac{p}{1-p}. \tag{9}$$

The results in **Table 12** show that the four variables MXDP, MXDV, NIMP, and RSPD were the most influential factors in generating abnormal values. Moreover, the values of the odds ratio are larger for these variables in the given order. This suggests that the probability of abnormal irregularity also increases the order of these variables. Among these four variables, MXDP and MXDV are related to LLID, whereas NIMP and RSPD are related to maintenance and train operation, respectively. It is interesting to note that these factors are more influential in causing abnormal LLID lots than other factors such as rail structural characteristics and facilities. Thus, we conclude that variables such as STDV, MXDV, NIMP, and RSPD are appropriate for inclusion in the LRM. In addition, we can confirm from omnibus testing that all variables are significant with significance probabilities of less than 0.001, which is much less than 0.05.

5.2. Applying Ordinal LRM Analysis for Finding Variation Factors in Lots to Be Maintained

We applied the ordinal LRM to the clusters {C1, C2, and C3} with a large LLID for the 6, 27, and 19 lots, respectively, in each of those clusters, which are the sets of lots to be maintained, as obtained by the cluster analysis. The dependent variables take values of 1, 2, and 3 for C1, C2, and C3, respectively, for the respective lots included in C1, C2, and C3, respectively. As mentioned in Section 5.1, nine independent variables were: MXDP, STDV, MXDV, NIMP, NRAD, BLST, NTRB, RSPD, and PTWT. These variables were considered to have a high probability of generating abnormal LLID occurrence in this LRM analysis. Parameter estimation results for ordinal LRM are shown in **Table 13**. We again attempted to reduce the variables to the regression equation, and each variable became significant using the variable reduction method. Because C1, C2, and C3 are all in the linear section, we excluded the variable CRVR. By applying the variable selection method while reducing variables so that the significance probability for each variable is less than 0.5, and deleting a variable one by one, we obtained numerical results with three variables, namely, MXDV, NIMP, and NRAD. The significance probabilities for all three variables MXDV, NIMP, and NRAD become less than 0.5, that is, they prove to be significant. The results are shown in **Table 14**, which indicates that the variables NIMP, NRAD, and MXDV are particularly significant in the given order as influencing factors that cause abnormal LLID occurrence. The fact that the odds ratio values are large in this order also indicates that the influence of these variables is highly related to the probability

Table 13. Parameter estimation results for ordinal LRM (9 variables).

Variable	z	Standard error	Wald square	Significance probability
MXDP	−1.083	0.835	1.681	0.195
STDV	−0.722	1.851	0.152	0.697
MXDV	−0.612	0.908	0.455	0.500
NIMP	0.302	0.155	3.783	0.052
NRAD	−1.493	0.524	8.123	0.004
BLST	−17.816	0		
NTRB	0.479	1.069	0.201	0.654
RSPD	0.004	0.017	0.053	0.818
PTWT	0.000	0	0.384	0.535

Table 14. Parameter estimation results for ordinal logit model (3 variables).

Variable	z	Standard error	Wald square	Significance probability	Exp(B)	Odds Ratio
MXDV	−1.629	0.406	16.06	<0.001	0.196	0.196
NIMP	0.274	0.147	3.50	0.061	1.31	1.31
NRAD	−1.330	0.460	8.35	0.004	0.265	0.265

of abnormal LLID occurrence. Among the three variables, NIMP and NRAD are related to maintenance management, and MXDV is a data-related variable for LLID. This suggests that factors based on maintenance-related indicators have a greater influence on abnormal LLID occurrence than other factors, such as LLID indicators related to rail structural characteristics, facilities, and equipment.

It is interesting to note that both the binomial and ordinal logit models showed similar results to the significant variables. This suggests that these variables are the most important factors in the occurrence, rapid advancement, and detection of abnormal LLID.

6. Summary and Conclusions

In this study, we focused on the railway business as a major social infrastructure that plays a fundamental role in transportation and logistics in many countries worldwide. We proposed and verified an efficient method for detecting lots that indicate abnormal LLID from the perspective of track maintenance management, which is necessary to enhance the reliability and safety of the railway system. We applied various types of multivariate analysis techniques such as cluster analysis, PCA, and binomial and ordinal LRMs to investigate the degree of abnormalities of the LLIDs. We can say that this approach is not only new, but it can lead to future important work focusing on RAMS (Reliability, Availability, Maintainability, Safety) and LCCA (Life Cycle Cost Analysis). We have been working to take the risk measurement into consideration (refer to [17]) and to investigate the serious train accidents in Japan using a probabilistic mathematical modeling approach (refer to [65]). These papers aim for improving reliability (R), maintainability (M), and safety (S). Our future problem exists in attempting to incorporate the factors related to RAMS-LCCA on the railway track elements into our OTMS model.

The results obtained from this study are summarized as follows.

1) We defined ten variables that influence the abnormal LLID lots, which can be mainly classified into three categories as follows. (i) Displacement-related: MXDP, STDV, MXDV, NIMP, NRAV; (ii) rail structure-related: Turnout, BLST, CRSL EXJT, CRVR, GIJT, and RGRL; (iii) train operation-related: RSPD and PTWT.

2) When the cluster analysis technique was applied using five variables to the sample set of lots, they were classified into five clusters, C1, C2, C3, C4, and C5, in the decreasing order from the one with the largest LLID to the one with the smallest. They were divided into two groups: {C1, C2, C3} and {C4, C5}.

3) The lots in clusters {C1, C2, C3}, accounting for approximately 10% of the total lots, required maintenance, whereas those in the remaining clusters {C4, C5}, accounting for approximately 90% of the total, showed good track condition and did not require maintenance.

4) Clusters {C1, C2, C3} generally contain lots with large values for MXDP, and NIMP exhibits large fluctuations in LLID and a large number of improvements. These include lots with particularly poor track conditions among the clusters. The lots in C4 are stable with no significant LLID and low maintenance frequency. C5, which contains the majority of lots, is classified as lots that have remained in a state of almost no fluctuation with an LLID close to 0 mm.

5) To analyze the background factors related to the detection of abnormal LLID lots, PCA was applied using data based on 14 variables. We assumed two cases: using data covering all 500 lots (Case 1) and using only 52 lots in {C1, C2, C3} with poor track conditions obtained by cluster analysis (Case 2). Regarding PC1, variables MXDV, MXDP, and STDV had large positive values, whereas variables NTBD and PTWT had small or negative values in Case 1, and variables EXJT, NIMP, and PTWT had either small or negative values in Case 2. Thus, we concluded that PC1 represents the axis from the values related to LLID data to those related to rail structure, equipment, and operation. The variables that had large positive values in PC2 in Cases 1 and 2 were BLST, PTWT, RGRL, and BRIN, while negative variables with large absolute values were NTBD and RSPD. Thus, we concluded that PC2 represents the axis from values related to rail structure and operation to those related to maintenance and operation.

6) Binomial LRM analysis for investigating the causes of the abnormal LLID lots revealed that by gradually eliminating the less influential variables from the 10 variables, namely, MXDP, STDV, MXDV, NIMP, NRAV, BLST, NTRB, CRVR, RSPD and PTWT, MXDP, MXDV, NIMP and RSPD were found to be the four most influential variables. Furthermore, the values of odds ratio were found to be larger for these variables in the given order, thus increasing the

probability of abnormal LLID.

7) Ordinal LRM was applied to the LLID of 6, 27, and 19 lots in each of the clusters {C1, C2, C3}, from 9 variables MXDP, STDV, MXDV, NIMP, NRAV, BLST, NTRB, RSPD and PTWT, using the variable reduction method. As a result of decreasing the variables until the regression equation and until each variable became significant, it was found that the three variables MXDV, NIMP, and NRAD were particularly influential in generating abnormal LLID lots. Furthermore, the odds ratio were found to be larger for these variables in the given order, thus increasing the probability of abnormal LLID.

We applied multivariate analysis techniques to validate our solution for selecting abnormal LLIDs. How to find abnormal LLIDs more quickly, more efficiently and more accurately to incorporate those solutions into our optimization OTMS model is one of the remaining problems we have been challenging. In addition, investigating the deterioration process of tracks in more detail, then obtaining more accurate forecasting explicitly is another remaining problem. In this study, we proposed and validated a method for efficiently detecting abnormal LLID lots, from the viewpoint of efficient track maintenance. Based on the results obtained here, a more efficient track maintenance management system can be constructed by providing data on the lots to be maintained to the optimal track maintenance planning and scheduling system, which we have developed and verified. Our system for optimal track maintenance planning and scheduling is already being used by the major railway companies in Japan. We aim to introduce the results of this study to these systems to construct a more efficient track maintenance management system.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] MLIT (Ministry of Land, Infrastructure, Transport and Tourism), Supervised by Railway Bureau of the Ministry of Land, Infrastructure and Transport (2020) Annual Report on Rail Statistics, Government Data Research Center of Japan. https://www.mlit.go.jp/tetudo/tetudo_tk6_000032.html

[2] NRI (Nomura Research Institute) (2022) What Is "Sustainable Regional Public Transportation" Considering the Corona Disaster and Population Decline? The 329th NRI Media Forum. (In Japanese)

[3] Ieda, H., Kanayama, Y., Ota, M., Yamazaki, T. and Okamura, T. (2001) How Can the Quality of Rail Services in Tokyo Be Further Improved? *Transport Policy*, **8**, 97-106. https://doi.org/10.1016/S0967-070X(01)00002-6

[4] Akita, T. (1997) Safety Is the Beginning of Railway Utilities. Railway Research Review, Railway Technical Research Institute, 10-11.

[5] Oyama, T. and Miwa, M. (2008) Considering the Safety Based on the Serious Train Accidents and Natural Disaster Damages Data in Japan. *Communications of Operations Research*, *Operations Research Society of Japan*, **53**, 11-17. (In Japanese)

[6] Tyrell, D.C. (2002) Rail Passenger Equipment Accidents and the Evaluation of Crashworthiness Strategies. *Proceedings of the Institution of Mechanical Engineers*, **216**, 131-147. https://doi.org/10.1243/09544090260082371

[7] Jamie, S. and Alan, N.B. (2004) British Railway Accident: A Systemic Analysis. *Railway Engineering* 2004, London, 2004.

[8] Huang, S. (2002) Safety Management in the UK Railway Network. *Railway Engineering* 2002, London, 2002.

[9] An, M. and Lupton, J. (2002) Safety Assessment in Railway—The Current Status and Future Aspects. *Railway Engineering* 2002, London, 2002.

[10] Fukuda, H. (1990) Problems in Railway Accidents Analysis and Safety Assessment Indices. *Quarterly Report of RTRI*, **31**, 2.

[11] Fukuda, H. (1997) Basic Study on the Safety Management System Methodology and Building Safety Management Support System. *Quarterly Report of RTRI*, **11**, 1-6.

[12] Miwa, M., Gozun, B. and Oyama, T. (2006) Statistical Data Analyses to Elucidate the Causes and Improve the Countermeasures for Preventing Train Accidents in Japan. *International Transactions in Operational Research*, **13**, 229-251. https://doi.org/10.1111/j.1475-3995.2006.00546.x

[13] Miwa, M. and Oyama, T. (2008) Considering Safety on the Serious Train Accidents in Japan. *Communications of Operations Research*, *Operations Research Society of Japan*, **53**, 11-17. (In Japanese)

[14] Yamaguchi, T., Miwa, M. and Oyama, T. (2015) Investigating the Occurrences of Train Accidents and Their Effects to the Train Operation in Japan. *Transportation Statistics*, *Research Institute for Transportation Statistics*, No. 41, 6-16. (In Japanese)

[15] Miwa, M. and Oyama, T. (2010) Proposal and Evaluation of Countermeasures for Preventing Serious Train Accidents by Analyzing Statistical Accident Data in Japan. *Journal of Civil Engineering Society D*, **66**, 89-105. (In Japanese)

[16] Miwa, M. and Oyama, T. (2011) Statistical Data Analysis for Damages Due to Train Accidents in Japan. *Proceeding of Japan-China Joint Workshop on Accident Prevention and Disaster Mitigation Policy*, Tokyo, 29 July 2011, 101-112. https://doi.org/10.2208/jscejd.66.89

[17] Miwa, M. and Oyama, T. (2018) An Optimal Track Maintenance Scheduling Model Analysis Taking the Risk of Accidents into Consideration. *International Transactions in Operational Research*, **25**, 1465-1490. https://doi.org/10.1111/itor.12425

[18] Miwa, M. and Mizuno, M. (2014) Modeling an Optimal Track Maintenance Schedule in Consideration of Train Derailment Risk. *Quarterly Report of RTRI*, **55**, 51-59. https://doi.org/10.2219/rtriqr.55.51

[19] Evans A.W, (2010) Rail Safety and Rail Privatization in Japan. *Accident Analysis and Prevention*, **42**, 1296-1301. https://doi.org/10.1016/j.aap.2010.02.007

[20] Evans, A.W. (2013) The Economics of Railway Safety. *Research in Transportation Economics*, **43**, 137-147. https://doi.org/10.1016/j.retrec.2012.12.003

[21] Evans, A.W. (2002) Speed and Rolling Stock of Trains in Fatal Accidents on Britain's Main Line Railways: 1967-2000. *Proceeding of the Institution of Mechanical Engineers*, **216**, 81-95. https://doi.org/10.1243/09544090260082326

[22] Evans, A.W. (2003) Estimating Transport Facility Risk from Past Accident Data. *Accident Analysis and Prevention*, **35**, 459. https://doi.org/10.1016/S0001-4575(02)00024-6

[23] Zhao, J., Chan, A.H.C. and Stirling, A.B. (2006) Risk Analysis of Derailment Induced by Rail Breaks—A Probabilistic Approach. *RAMS'06*. *Annual Reliability and*

Maintainability Symposium, Newport Beach, 23-26 January 2006, 486-491.

[24] Zhao, J., Chan, A. and Burrow, M. (2007) Reliability Analysis and Maintenance Decision for Railway Sleepers Using Track Condition Information. *Journal of the Operational Research Society*, **58**, 1047-1055. https://doi.org/10.1057/palgrave.jors.2602251

[25] Cesario, P., Sacco,N. and M. Sciutto, M. (2008) A Discrete Time Markov Chain Approach to Global Risk Analysis in Railway Transportation. *Computers in Railways*, **11**, 123-131. https://doi.org/10.2495/CR080131

[26] El-Koursi, E.-M. and Bruyelle, J.-L. (2016) Railway Accident Prevention and Infrastructure Protection. *Journal of Civil Engineering and Architecture*, **10**, 96-107. https://doi.org/10.17265/1934-7359/2016.01.010

[27] Liang, C., Ghazel, M., Cazier, O., Bouillaut, L., Miloudi, E. and El Koursi, E. (2017) Bayesian Network Modeling Applied on Railway Level Crossing Safety. 2*nd International Conference Reliability*, *Safety and Security of Railway Systems*: *Modelling*, *Analysis*, *Verification and Certification*, Pistoia, November 2017, 15 p. https://hal.archives-ouvertes.fr/hal-01815714

[28] Bargegola, I., Gilani, N.M. and M. Abolfazlzadehc, M. (2017) Statistical Analysis of the Railway Accidents Causes in Iran. *International Journal of Engineering*, *Transactions C*: *Aspects*, **30**, 1822-1830. https://doi.org/10.5829/ije.2017.30.12c.02 https://www.researchgate.net/publication/322469986

[29] Bureika, G., Gaidamauskas, E., Kupinas, J., Bogdevičius, M. and Steišūnas, S. (2017) Modelling the Assessment of Traffic Risk at Level Crossings of Lithuanian Railways. *Transport*, **32**, 282-290. https://doi.org/10.3846/16484142.2016.1244114

[30] Hadj-Mabrouk, H. (2019) Contribution of Artificial Intelligence to Risk Assessment of Railway Accidents. *Urban Rail Transit*, **5**, 104-122. https://doi.org/10.1007/s40864-019-0102-3

[31] Fujita, K. (2017) Condition Based Maintenance for Japan's Railways. *Global Railway Review*, **10**. https://www.globalrailwayreview.com/article/61524/condition-based-maintenance-japan/

[32] Liden T. (2015) Railway Infrastructure Maintenance—A Survey of Planning Problems and Conducted Research. *Transportation Research Procedia*, **10**, 574-583. https://doi.org/10.1016/j.trpro.2015.09.011

[33] Sedghi, M., Kauppila, O., Bergquist, B., Vanhatalo, E. and Kulahci, M. (2021) A Taxonomy of Railway Track Maintenance Planning and Scheduling: A Review and Research Trends. *Reliability Engineering & System Safety*, **215**, Article ID: 107827. https://doi.org/10.1016/j.ress.2021.107827

[34] Budai, G., Huisman, D. and Dekker, R. (2005) Scheduling Preventive Railway Maintenance Activities. *Journal of the Operational Research Society*, **57**, 1035-1044. https://doi.org/10.1057/palgrave.jors.2602085

[35] Budai-Balke, Dekker, R. and Kaymak, U. (2009) Genetic and Memetic Algorithms for Scheduling Railway Maintenance Activities. Econometric Institute Report EI, 30.

[36] Miwa, M. and Oyama, T. (2004) All-Integer Type Linear Programming Model Analyses for the Optimal Railway Track Maintenance Scheduling. *OPSEARCH*, **41**, 155-164. https://doi.org/10.1007/BF03398841

[37] Oyama, T. and Miwa, M. (2006) Mathematical Modeling Analyses for Obtaining an Optimal Railway Track Maintenance Schedule. *Japan Journal of Industrial and Applied Mathematics*, **23**, 207-224. https://doi.org/10.1007/BF03167551

[38] Tanaka, H. and Miwa, M. (2019) Modeling the Development of Rail Corrugation to Schedule a More Economical Rail Grinding. *Proceedings of the Institution of Mechanical Engineers*, *Part F*: *Journal of Rail and Rapid Transit*, **234**, 370-380. https://doi.org/10.1177/0954409719894833

[39] Miwa, M., Nishijima, Y., Yasaka, K., Matsumoto, M., Yamada, H. and Oyama, T. (2021) Optimal Railway Track Maintenance Scheduling Model for Ballast Replacement Work by Considering Several Numbers of Maintenance Machines for the Work. *Journal of Japan Society of Civil Engineers*, *D3* (*Infrastructure Planning and Management*), **77**, 260-279. https://doi.org/10.2208/jscejipm.77.3_260

[40] Matsumoto, M. and Miwa, M. (2017) Modeling an Optimal Track Maintenance Schedule in Consideration of Timing of Grinding and Tamping. *Quarterly Report of RTRI*, **58**, 229-235. https://doi.org/10.2219/rtriqr.58.3_229

[41] Vale, C., Ribeiro, I.M. and Calgada, R. (2012) Integer Programming to Optimize Tamping in Railway Tracks as Preventive Maintenance. *Journal of Transportation Engineering*, **138**, 123-131. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000296

[42] Caetano, L.F. and Teixeira, P.F. (2016) Strategic Model to Optimize Railway-Track Renewal Operations at a Network Level. *Journal of Infrastructure Systems*, **22**, Article ID: 04016002. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000292

[43] Caetano, L.F. and Teixeira, P.F. (2016) Predictive Maintenance Model for Ballast Tamping. *Journal of Transportation Engineering*, **142**, Article ID: 04016006. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000825

[44] RTRI (Railway Technology Research Institute) (2007) Handbook of Train Structures Maintenance Managements (Track Maintenance). 9-110.

[45] BS EN 13231-1:2013 (2013) Railway Applications—Track. Acceptance of Works on Ballasted Track. Plain Line, Switches and Crossings, BSI Stand Publication.

[46] BS EN 13306 (2017) Maintenance—Maintenance Terminology. BSI Stand Publication.

[47] Meier-Hirmer, C., Senee, A, Gilles, R., Sourget, F. and Roussignol, M. (2006) A Decision Support System for Track Maintenance. *WIT Transactions on the Built Environment*, **88**, 217-226. https://doi.org/10.2495/CR060221

[48] Peng, F. (2011) Scheduling of Track Inspection and Maintenance Activities in Railroad Networks. Ph.D. Thesis, University of Illinois, Urbana.

[49] Xu, P., Liu, R., Sun, Q. and Wang, F. (2012) A Novel Short-Range Prediction Model for Railway Track Irregularity. Discrete Dynamics in Nature and Society, 1-12. https://doi.org/10.1155/2012/591490

[50] Bai, L., Liu, R., Sun, Q., Wang, F. and Wang, F. (2014) Classification-Learning-Based Framework for Predicting Railway Track Irregularities. *Proceedings of the Institution of Mechanical Engineers*, *Part F*: *Journal of Rail and Rapid Transit*, **230**, 598-610. https://doi.org/10.1177/0954409714552818

[51] Hu, C. and Liu, X. (2016) Modeling Track Geometry Degradation Using Support Vector Machine Technique. *ASME*/*IEEE Joint Rail Conference*, South Carolina, 12-15 April 2016, V001T01A011 (JRC2016-5739). https://doi.org/10.1115/JRC2016-5739

[52] Tsubokawa, Y., Yazawa, E., Ogiso, K. and Anmoku, T. (2012) Development of the Car Body Mounted Track Measuring Device with the Inertial Mid-Chord Offset Method. *RTRI Quarterly Report*, **53**, 216-222. https://doi.org/10.2219/rtriqr.53.216

[53] Tanaka, H., Matsumoto, M. and Harada, Y. (2016) Application of Axle-Box Acceleration to Track Condition Monitoring for Rail Corrugation Management. *7th*

IET Conference on Railway Condition Monitoring 2016 (*RCM* 2016), Birmingham, 27-28 September 2016, 1-7. https://doi.org/10.1049/cp.2016.1191

[54] Tanaka, H., Matsumoto, M., Miwa, M. and Miyazaki, Y. (2017) Comparison Analysis of Various Evaluation Indexes of Track Irregularity Data for High-Speed Railway Track. *Proceedings of Railway Engineering* 2017, Edinburgh, June 2017.

[55] Gibert, X., Patel, V.M. and Chellappa, R. (2017) Deep Multitask Learning for Railway Track Inspection. *IEEE Transactions on Intelligent Transportation Systems*, **18**, 153-164. https://doi.org/10.1109/TITS.2016.2568758

[56] Chia, L., Bhardwaj, B., Lu, P. and Bridgelall, R. (2019) Railroad Track Condition Monitoring Using Inertial Sensors and Digital Signal Processing: A Review. *IEEE Sensors Journal*, **19**, 1558-1748. https://doi.org/10.1109/JSEN.2018.2875600

[57] Jamshidi, A., Hajizadeh, S., Su, Z., Naeimi, M., Núñez, A., Dollevoet, R., Schutter, B. and Li, Z. (2018) A Decision Support Approach for Condition-Based Maintenance of Rails Based on Big Data Analysis. *Transportation Research Part C: Emerging Technologies*, **95**, 185-206. https://doi.org/10.1016/j.trc.2018.07.007

[58] Thiyagarajan, J. S., Su, D., Tanaka, H., Zhao, B. and Nagayama, T. (2021) Response Based Track Profile Estimation Using Observable Train Models with Numerical and Experimental Validations. *Smart Structures and Systems*, **27**, 267-284.

[59] Le, Y., Oka, M. and Kato, H. (2022) Efficiencies of the Urban Railway Lines Incorporating Financial Performance and In-Vehicle Congestion in the Tokyo Metropolitan Area. *Transport Policy*, **116**, 343-354. https://doi.org/10.1016/j.tranpol.2021.12.017

[60] Jituzumi, T. and Nakamura, A. (2006) Incentive-Compatible Subsidies for the Japanese Railway Sector. *Journal of Public Utility Economics*, **57**, 21-29.

[61] Chenariyan, M., Djoerd, N., Mariëlle, H., Stoelinga, M. and van Noort, M. (2019) The Recent Applications of Machine Learning in Rail Track Maintenance: A Survey. In: Collart-Dutilleul, S., *et al.*, Eds., *Reliability, Safety, and Security of Railway Systems. Modelling, Analysis, Verification, and Certification*, Springer, Berlin, 91-105.

[62] Sadeghi, J. and Askarinejad, H. (2012) Application of Neural Networks in Evaluation of Railway Track Quality Condition. *Journal of Mechanical Science and Technology*, **26**, 113-122. https://doi.org/10.1007/s12206-011-1016-5

[63] Lee, J.S., Hwang, S.H., Choi, I.Y. and Kim, I.K. (2018) Prediction of Track Deterioration Using Maintenance Data and Machine Learning Schemes. *Journal of Transportation Engineering Part A Systems*, **144**, Article ID: 4018045. https://doi.org/10.1061/JTEPBS.0000173

[64] Kimura, H. and Tanaka, H. (2009) Detection Methods for Finding Quickly Deteriorating Lots for Vertical Displacement Irregularities in Ballast Tracks. *Proceedings of the 16th Symposium on the Railway Technology Association*, Tokyo, October 2009, 277-280. (In Japanese)

[65] Oyama, T. and Miwa, M. (2022) Applying Probabilistic Mathematical Modeling Approach and AI Technique to Investigate Serious Train Accidents in Japan. *Sustainability Analytics and Modeling*, **2**, Article ID: 100005. https://doi.org/10.1016/j.samod.2022.100005