Track Geometry Defect Rectification Based on Track Deterioration Modeling and Derailment Risk Assessment

Abstract

Analyzing track geometry defects is critical for safe and effective railway transportation. Rectifying the appropriate number, types and combinations of geo-defects can effectively reduce the probability of derailments. In this paper, we propose an analytical framework to assist geo-defect rectification decision making. Our major contributions lie in formulating and integrating the following three data-driven models: 1) A track deterioration model to capture the degradation process of different types of geo-defects; 2) A survival model to assess the dynamic derailment risk as a function of track defect and traffic conditions; 3) An optimization model to plan track rectification activities with two different objectives: a cost-based formulation (CF) and a risk-based formulation (RF). We apply these approaches to solve the optimal rectification planning problem for a real-world railway application. We show that the proposed formulations are efficient as well as effective, as compared to existing strategies currently in practice.

Key words: Rail transportation; Track geometry defects; Track deterioration model; Track derailment risk; Track defect rectification
1. Introduction

Rail is a crucial mode of transportation in the United States. According to the National Transportation Statistics report from the Bureau of Transportation Statistics (Association of American Railroads 2011), 42.7% of the United States freight revenue ton-miles were carried by railroad; this represents the largest portion of the inter-city freight market.

Detection and rectification of track defects are major issues in the railway industry. The existing literature categorizes these defects into one of two groups: track structural defects and track geometry defects (Sadeghi & Askarinejad 2010). Track structural defects are generated from the structural conditions of the track, which include the condition of the rail, sleeper, fastening systems, subgrade and drainage systems. On the other hand, track geometry defects (referred to as geo-defects in the remainder of this paper) indicate severe ill-conditioned geometry parameters such as profile, alignment, gage, etc., as shown in Figure 1 (ARC-TECH.NET 2012).

Track defects have become the leading cause of train accidents in the United States since 2009. 658 of 1,890 (34.8%) train accidents were caused by track defects in 2009, incurring a $108.7 million loss (Peng 2011). Therefore, it is imperative to understand the deterioration process of railway track systems and facilitate track maintenance planning. Previous studies on track deterioration divide track segments into several shorter sections for analyzing summary statistics of raw geometry measurements (Federal Railroad Administration 2005; Sadeghi & Askarinejad 2010; El-Sibaie & Zhang 2004; Berawi et al. 2010). The overall statistics provide a measure of segment quality, called Track Quality Indices (TQIs). TQIs have been widely used for railway maintenance scheduling, since they provide a high-level assessment of railway track performance (Tolliver & Benson 2010; Association of American Railroads 2011). However, TQIs only provide an aggregate level picture and they cannot identify individual severe geo-defects for track rectification.

Maintaining the existing tracks through rectification and renewal is critical to railroad operation and safety. In 2008, Class I railroads, defined as line haul freight railroads with operating revenues of $398.7 million or more (Association of American Railroads 2011), spent $7.52 billion on track maintenance (Tolliver & Benson 2010). Track maintenance activities can be categorized into two main groups: preventive maintenance and corrective maintenance (Andersson 2002). Preventive maintenance is pre-planned and carried out to avoid future defects, whereas corrective maintenance rectifies existing defects in the infrastructure. Most of the literature in this area describes preventive or planned maintenance (Beichelt & Fischer 1980; Higgins 1998; Cheung et al. 1999; Higgins et al. 1999; Simson et
al. 2000; Budai et al. 2006; Oyama & Miwa 2006; van Zante–de Fokkert et al. 2007; Peng et al. 2011) due to its large scale of operation and high complexity, whereas very few studies have addressed the problems in corrective or unplanned maintenance (Zhao et al. 2007), referred to as track rectification (or track repair) in this paper.

Track rectification is subsumed by the broader area of railway project selection and planning, which uses prioritization and optimization techniques. Prioritization is essentially performed in a sequential manner by first enlisting the maintenance projects required to be executed. Once projects are identified, the next step is to prioritize the projects based on their relative perceived urgency. Projects with the highest priority are executed until all finances are expended. Remaining projects are re-prioritized together with the new projects upon availability of funds.

A popular multi-criteria decision making methodology for prioritization is the Analytic Hierarchy Process (AHP), first presented by Saaty (1980). After interviewing six track managers, Nyström and Söderholm (2010) applied AHP techniques to understand different criteria from decision-makers and then rank maintenance actions. Cheng et al. (2012) combined the AHP with artificial neural network (ANN) methodologies to formulate a ranking predication model by connecting objective measures with subjective judgments. However, the AHP suffers from a number of drawbacks and is often criticized for the way the criteria weights are elicited, for rank reversal problems, and for fundamental issues with the theory (Boucher et al. 1997). Typically, prioritization involves ranking rectification activities based on the “worst first” principle, and therefore fails to account for the change in benefit for the funds expended; it may produce a maintenance strategy which could be far from optimal.

Unlike prioritization approaches, optimization-based methods have the ability to develop a maintenance plan while capturing the effect of deferring a maintenance activity on the track condition. An ideal optimization approach is one that evaluates all possible repair strategies at the network level without imposing unnecessary constraints or subjective judgments. Niemeier et al. (1995) constructed five optimization models, each building on a basic linear-programming formulation, for selecting an optimal subset of projects submitted for a statewide programming process. Ahern and Anandarajah (2007) developed a weighted integer goal-programming model for selecting railway projects for investment while maximising the objectives and meeting the budget limit for capital investment. In this paper, we extend the literature on optimization-based methods by proposing a data-driven formulation that incorporates track deterioration and derailment risk models.
2. Background and Model Preliminaries

2.1. Background

Track rectification decisions are typically made by the local track master in the network. A local track master is responsible for the railway infrastructure in a certain area, containing one or several track sections, typically from dozens of miles to a couple of hundred miles (Nyström & Söderholm 2010). Among other responsibilities, s/he is responsible for rectifying track geometry defects, thereby ensuring high standards of safety and cost effectiveness.

According to the US Federal Railroad Administration (FRA) track safety standards, individual defects whose amplitudes exceed a certain tolerance level must be treated properly. Traditionally, geometry cars classify each defect into two severity levels, denoted in this paper as either “Red tags” or “Yellow tags”. Red tags are defects that are in violation of FRA track safety standards, and railroads must fix these defects as soon as possible after their discovery or else they risk being fined. Yellow tags are defects whose amplitudes are currently below FRA limits, and they may or may not meet the particular railroad's own standards for rectification. According to current practice, railroads fix Red tags within a due date after inspection and they examine the Yellow tags, repairing very few of them based on their severity as estimated from field experience. Whether to fix a Yellow tag or not may depend on several factors, such as the state of the track geometry, defect history, curvature of the track, MGT (Million Gross Tons), consequential derailment cost, etc. It may be particularly prudent to rectify extreme severe Yellow tag geo-defects that are likely to soon become Red tag defects. To our best knowledge, there is not much prior literature on the individual geo-defect rectification process.

2.2. Data Summary

Our models are based on field datasets from a Class-I railroad, including 3-year traffic data, derailment data, and geo-defect data from January 2009 to December 2011. Since main line tracks carry most of the traffic, and derailments associated with these tracks usually cost much more than other track types, we focus our analysis on about 2000 miles of main line tracks in this study. In total, there are approximately 4,000 Red tag defects and 27,000 Yellow tag defects. The dataset contains around 40 different types of geo-defects. The top 12 major geo-defect types pertaining to our analysis are described in Table 1.

For modeling purposes, the datasets are processed along both spatial and temporal dimensions:
• Spatially, the rail network is defined by line segments (they usually connect two cities), track numbers (0-8 for main line tracks) and mile post locations. Constructed in such a fashion, the rail lines range from a few miles to hundreds of miles. To generate consistent spatial units and accommodate different modeling purposes, we divide the main line network further into two different levels of smaller segments, called *lots* and *sections*. Each lot is 0.02 mile (about 100ft) in length, used for track deterioration analysis. At a higher level, a continuous track segment is divided into 2 mile long sections, used for track derailment risk and geo-defect rectification modeling.

• Temporally, regular track geometry inspection is performed 1 to 4 times per year according to characteristics of each track segment. Geo-defects are reported and updated after each inspection run. When they occur in the same inspection run window, different types of geo-defects are aggregated to the level of an inspection run, mainly for derailment risk analysis.

### 2.3. Model Summary

In order to improve current track rectification decisions, this study aims to help existing railroads address the following three questions: 1) How Yellow tags of each type deteriorate into Red tags; 2) How unrectified Yellow tags affect derailment risk; 3) How to prioritize and rectify Yellow tags within a limited repair horizon. The main objective of this study is to propose a framework for making optimal track geo-defect rectification decisions, in order to appropriately reduce the probability of a derailment as well as its associated costs. Additionally, effective track geometry maintenance reduces dynamic vehicle and track interaction, thus reducing the stress state of the railroad. Compared with the existing state-of-the-practice decision model depicted in Figure 2, our major contributions lie in formulating and integrating the following three models:

- A track deterioration model to capture the degradation process of different types of geo-defects.
- A survival model to assess the dynamic derailment risk as a function of the current track and traffic condition.
- An optimization model to plan track rectification activities that either minimizes the highest derailment risk or the total expected costs, including potential derailment costs and rectification costs.
To make the optimization model tractable, we make the following assumptions:

- **Assumption 1**: Red tags may be rectified individually, but Yellow tags are rectified in bulk, at the section and defect type level. This means that the decision of rectifying yellow tags are made for each defect type in each section level.

- **Assumption 2**: The repair horizon, from a few days up to a few weeks, is much smaller than the track inspection interval, typically ranging from 3 months to 1 year. Therefore, once the repair activity is determined, we assume that derailment risk will take effect immediately according to the repair activities scheduled.

- **Assumption 3**: The travel time between two defects is negligible as compared with the time spent at the defect rectification location.

- **Assumption 4**: Defects to be repaired at the same time step constitute one cluster. The center of each cluster is defined as the defect with median post mile. The total distance travelled is approximated by the total distance from the defects to their cluster centers.

3. The Track Deterioration Model

We develop a statistical track deterioration model for Yellow tag defects, representing the causes and consequences of track deterioration. The model takes various factors into account, including the current track conditions and traffic information, and has the capability to predict future track conditions. The track deterioration process is captured by studying geo-defect amplitude changes, measured during each geometry inspection run. The statistical model constructs the relationships between the effective parameters and the track deterioration rate, while incorporating uncertainty arising from environmental factors and measurement noise. The statistical model is able to predict the deterioration of each geo-defect and the probability of whether a Yellow tag geo-defect will become a Red tag within a given duration.

To model track deterioration, we track the evolution of track defects. However, due to the lack of geo-defect indices, it is not possible to track any particular geo-defect over time. We handle this situation by tracking the condition of small track segments, where each segment contains very few geo-defects for each inspection run. First we divide the tracks into non-overlapping lots of equal length 0.02 miles (105.6 feet). Then we aggregate the defects within each lot by inspection run for each defect type. We take the 90 percentile of the amplitudes to represent the track segment condition for the inspection run under consideration.
Exploratory analysis suggests fitting different models for different defect types, since the model parameters have varying effects on deterioration rate for each defect type. For example, GAGE_W1 and GAGE_W2 (see Table 1) exhibit different deterioration rates with respect to traffic in MGT. We assume that geo-defects get worse over time, i.e., defect amplitudes increase when there is no maintenance work. For each defect type, let \( y_k(t) \) denote the aggregated geo-defect amplitude (the 90 percentile of the defect amplitudes) of the track lot \( k \) at inspection time \( t \). The deterioration rate or the amplitude change rate over time \( \Delta t \) can be represented by \( (y_k(t + \Delta t) - y_k(t))/\Delta t \). We model the deterioration rate (for any given defect type) as follows:

\[
\log\left(\frac{y_k(t + \Delta t) - y_k(t)}{\Delta y_k(t)}\right) = \alpha_0 + \alpha_1 X_{1k}(t) + \cdots + \alpha_p X_{pk}(t) + \varepsilon_k(t) \quad \forall k = 1...N \quad (1a)
\]

where \( N \) is the total number of track lots. \( X_{pk}(t) \) is the \( p \)th external factor or predictor for \( k \)th track lot at inspection time \( t \). Based on our exploratory data analysis, we find that the distribution of \( (y_k(t + \Delta t) - y_k(t))/\Delta y_k(t) \) is highly skewed and becomes close to normal after making a log transformation. Therefore, we choose to use an exponential relationship between the external factors or predictors \( X_{1k}(t), \ldots, X_{pk}(t) \) and the deterioration rate in our model, similar to Sadeghi and Askarinejad (2010). The random error \( \varepsilon_k(t) \) is assumed normally distributed with mean 0 and standard deviation \( \sigma^2 \).

The factors included in our model are: monthly traffic MGT traveling through track lot \( k \) \( (X_{1k}(t)) \), monthly total number of cars \( (X_{2k}(t)) \), monthly total number of trains \( (X_{3k}(t)) \) and number of inspection runs in sequence since the last observed Red tag geo-defect\((X_{4k}(t))\). Model fitting shows that the factors have different impacts on deterioration rates for each defect type. The estimated coefficients, \( \alpha_0, \alpha_1, \ldots, \alpha_p \), are presented in Table 2, where \( \alpha_0 \) depicts the intercept of the model, and \( \alpha_i \) represents the coefficient for \( i \)th \( X \) factor. Mean squared errors (MSE), as shown in Table 2, are used for model selection: A model is chosen based on minimizing MSE. We test different models with different nonlinear functions of the factors, \( X_{pk}(t) \). The model in \((1a)\) has the smallest MSE.

Table 2 confirms that most defects deteriorate faster when the traffic load (either MGT, number of cars or number of trains) increases (Sadeghi & Askarinejad 2010). Contrary to traditional deterioration models for aggregated track quality index, our proposed model aims to predict the deterioration rate for each individual Yellow tag defect. It is therefore important
to leverage the Yellow tag existing duration, indicated by the sequence number of inspections since the last Red tag ($\alpha_d$). The estimated set of coefficients for $\alpha_d$ also demonstrates that Yellow tags deteriorate at an increasing rate. Due to limited data, some defect types, such as ALIGN, HARM_X, REV_X and SUPER_X, do not have any significant predictive factors in the model. This means that the deterioration rates for these defect types only depend on the current amplitude and not on traffic and other factors.

To compute the probability of a Yellow tag defect becoming a Red tag in the future, we predict the defect amplitude for the next inspection run according to equation 1(a). Based on current practice in geometry inspection, we choose $\Delta t$ as 90 days. Let the threshold for a Yellow tag defect becoming a Red tag for a certain defect type be $r$. Assuming that both $r$ and current amplitude $y_k(t)$ are positive, and that $y_k(t)$ is less than $r$, we define

$$h_k(t) = \log \left( \frac{r - y_k(t)}{\Delta y_k(t)} \right)$$

as the log-transformation of the deterioration rate threshold at current amplitude $y_k(t)$. According to the proposed model in (1a), the log-transformation of deterioration rate is assumed to be normally distributed. We can calculate the probability of a Yellow tag at time $t$ on track $k$ becoming a Red tag in $\Delta t$, by computing the upper quantile of the transformed quantity, $z = \log \left( \frac{y_k(t + \Delta t) - y_k(t)}{\Delta y_k(t)} \right)$, which follows a normal distribution. Then the probability $P^R_k(t)$ of a Yellow tag geo-defect at time $t$ on track lot $k$ becoming a Red tag in $\Delta t$ is

$$P^R_k(t) = \int_{h_k(t)}^{\infty} zdz \quad (1b)$$

### 4. The Track Derailment Risk Model

State-of-practice track geometry analysis and risk estimation systems mainly focus on current static track derailment risk utilizing mechanic models, such as ZETA-TECH’s TrackSafe model (Bonaventura et al. 2005) and TTCI’s Performance Based Track Geometry (PBTG) model (Li et al. 2004). Instead, this study analyzes large amounts of historical data and predicts the future derailment risk given both traffic and defect conditions.

Survival analysis is the field of study dealing with the analysis of data regarding the occurrence of a particular event, within a time period after a well-defined time origin (Collett 2003). Analyzing survival times is common in many areas, for instance, in biomedical
computation, engineering and the social sciences. In our railway application, each inspection run will “refresh” the track segment since all Red tag geo-defects will be repaired. If there is no derailment between two scheduled inspection runs on a track segment, the track can be considered to have “survived” from one inspection to the next. If any derailment occurs, the track segment is said to have failed in the time period since the last inspection run.

We refer to derailment on a particular track section as a hazard. In survival theory, there are three basic functions: the density function \( f(t) \), survival function \( S(t) \) and hazard function \( \lambda(t) \). For a derailment, density function \( f(t) \) expresses the likelihood that the derailment will occur at time \( t \). The survival function represents the probability that the track section will survive until time \( t \):

\[
S(t) = \text{Prob}(T \geq t) = \int_t^\infty f(x)dx = 1 - \int_0^t f(x)dx = 1 - F(t)
\]  

(2a)

where \( T \) denotes the survival time of a track segment after inspection and \( F(t) \) denotes the cumulative function of variable \( T \). The hazard function represents the instantaneous rate of failure probability at time \( t \), given the condition that the event has survived to time \( t \). By definition, the relationships between these three functions are:

\[
\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d\ln S(t)}{dt}
\]

\[
S(t) = \exp[-\int_0^t \lambda(x)dx]
\]

\[
f(t) = \lambda(t)S(t) = \lambda(t)\exp[-\int_0^t \lambda(x)dx]
\]

Parametric models such as the exponential, Weibull, log-logistic, or log-normal distributions may be used to specify the density distribution \( f(t) \), but such pre-defined distributions may be inappropriate for real world data. Without having to specify any assumptions about the shape of the baseline function, Cox (1972) proposed a semi-parametric method for estimating the coefficients of covariates in the model using the method of partial likelihood (PL) rather than maximum likelihood. This model assumes that the covariates multiplicatively shift the baseline hazard function and is by far the most popular choice in practice due to its elegance and computational feasibility (Cleves et al. 2004). Furthermore, unlike non-parametric analysis such as the Kaplan-Meier method and the rank test, it allows both nominal and continuous variables. The hazard function form of the Cox model is:
\[
\lambda(t; \beta, X) = \lambda_0(t)e^{\beta'X} \tag{2b}
\]

where \(\lambda_0(t)\) is an unspecified nonnegative function of time called the baseline hazard, \(\beta\) is a column vector of coefficients to be estimated, and \(\beta'X = \beta_0 + \beta_1x_1 + \beta_2x_2 + ... + \beta_kx_k\). Since the hazard ratio for two subjects with fixed covariate vectors \(X_i\) and \(X_j\),

\[
\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t)e^{\beta'X_i}}{\lambda_0(t)e^{\beta'X_j}},
\]

is constant over time, the model is also known as the proportional hazards (PH) model. In order to estimate \(\beta\), Cox (1972) proposed a conditional (or partial) likelihood function which depends only on the parameter of interest, proving that the resulting parameter estimators from the partial likelihood function would have the same distributional properties as full maximum likelihood estimators (Cox 1975). The partial likelihood function is described as

\[
L_p(\beta) = \prod_{i=1}^{n} \left[ \frac{e^{\beta'X_i}}{\sum_{j\in R(t_i)} e^{\beta'X_j}} \right]^{\delta_i},
\]

and the maximum partial likelihood estimator is found by solving the equation,

\[
\frac{\partial \ln(L_p(\beta))}{\partial \beta} = 0.
\]

One may refer to (Therneau & Grambsch 2000) for details of implementing a Cox model using existing statistical software. As described in Section 2, all the raw geo-defects are spatially aggregated to the section level (2 mile), and temporally into each inspection level. In any particular aggregated record, the dependent variable is either the time duration between two inspection runs (the censored survival time), or time duration between the derailment and the last inspection run before derailment. Selected candidate predictors are listed as follows:

- Monthly traffic in MGT
- Number of Yellow tag geo-defects (starting with “numYEL” in Table 3) in each defect category
- 90 percentile amplitude (starting with “amp90” in Table 3) of Yellow tag geo-defects in each defect category

The final Cox model fit to the censoring derailment data is illustrated in Table 3. An efficient way to evaluate the fitted model is to use Cox-Snell residuals (Cox & Snell 1968),

\[
r_{CS} = e^{\beta'X_i}\hat{H}_0(t_i),
\]

where \(\hat{H}_0(t_i)\) is the empirical cumulative hazard function evaluated at time \(t_i\).
where $\hat{H}_0(t_i)$ is the estimated integrated baseline hazard (or cumulative hazard). These residuals are based on the observation that, for a random time to event $T$ with survivor function $S(T)$, the random variable $-\log S(T)$ is distributed exponential with mean one. If the model is calibrated correctly, the Cox-Snell residuals should show a standard exponential distribution with hazard function equal to one, and thus the cumulative hazard of the Cox-Snell residuals should follow a straight 45 degree line. The plot in Figure 3 confirms that most of the step lines are close to the dashed straight line, except for a few tail large ones. As a result, we feel there is no evidence to reject the model.

The model shown in Table 3 includes 10 simple covariates, where each significant covariate represents a particular geo-defect type. A positive coefficient implies that the hazard is higher (hazard ratio is greater than 1.0), whereas a negative one indicates a lower hazard (hazard ratio is less than 1.0). In the fitted model, all covariates have positive coefficients, indicating that all geo-defect types listed in Table 3 have a strong positive impact on derailment risk. The higher the values of the covariates are, the higher is the derailment risk. Initially, geo-defect data of each type is aggregated into two types of covariates: the number based and amplitude based covariates. The number based covariates count the number of defects in each section for each type, whereas the amplitude based covariates calculate 90 percentiles of geo-defect amplitudes. After careful model selection, we restrict the number based group to GAGE_W1, REV_X and GAGE_W2; on the other hand, the amplitude based group includes SUPER_X, GAGE_C, DIP, WARP, HARM_X, WEAR, and ALIGN.

Recall from equation 2(a) that we can derive the derailment probability $P_{D_i}^O(t)$ on section $i$ prior to time $t$ as:

$$P_{D_i}^O(t) = \text{Prob}_i(T \leq t) = 1 - S_i(t),$$

(2c)

where $S_i(t)$ indicates the survival probability on section $i$ prior to time $t$. After some defects are rectified, the derailment probability will decrease, based on modifications to the hazard function $\lambda_i(t)$ on section $i$ at time $t$, as determined by equation 2(b). (The covariate vector $X$ is specified in Table 3). To accommodate different rectification alternatives (or activities) $a$, we denote $\lambda_{ia}(t)$ and $S_{ia}(t)$ as the hazard function and survival probability with respect to time $t$, respectively, after rectification activity $a$ is performed on section $i$. Each rectification activity corresponds to fixing a single type or combination types of geo-defects, equivalent to setting zeros for the relevant covariates. The derailment probability after rectification activity $a$ is taken, $P_{ia}^O(t)$, is therefore:
Given both the Yellow tag deterioration probability $P^R_K(t)$ and the track derailment probability $P^D_{ia}(t)$, we are ready to formulate the track rectification optimization model. When the rectification problem involves repairing a large amount of Yellow tags, it makes the problem difficult to solve. To reduce the computational complexity, we assume that although rectification decisions are made for individual defects with Red tags, multiple Yellow tags are rectified together at the section and defect type level. In summary, the proposed optimization model aims to optimize the timing to rectify Red tags within their repair due dates, whereas optimize both selection and rectification timing for Yellow tags before they deteriorate to Red. Defects that are to be repaired in the same time step constitute a cluster. The center of each cluster is defined as the defect with median post mile.

Let $\Delta t$ be the time interval between two inspection runs, and let $T_R$ be the total number of discrete rectification periods (e.g. days, weeks). The model parameters are as follows:

- $i \in I$: Index for track sections
- $a \in A$: Index for rectification activities
- $k, l \in K_a$: Indices for all defects, including both Yellow and Red tags
- $k' \in K_y$: Index for Yellow tags, $K_y \subseteq K_a$
- $t \in \{1, ..., T_R\}$: Index for defect rectification periods
- $L_k$: Due period to repair defect $k$
- $\tau_k$: Repair time for defect $k$
- $C_k$: Repair cost for defect $k$
- $C_{kl}$: Travel cost from defect $k$ to $l$
- $C_D$: Derailment cost
- $P^D_{ia}(\Delta t)$: Probability of a derailment in the inspection time interval $\Delta t$, if alternative $a$ is chosen for section $i$
- $P^R_K(\Delta t)$: Probability that Yellow tag defect $k'$ will progress to a Red tag defect in the inspection time interval $\Delta t$, if it is not chosen to be rectified
- $\lambda_k$: If $\lambda_k = 1$, defect $k$ has to be rectified in horizon $T_R$, including all the red tags and yellow tags with $P^R_K(\Delta t) \geq \delta$, where $\delta$ represents predefined probability; If $\lambda_k = 0$, we have the flexibility to decide whether to rectify defect $k$ or not
- $\mu_{iak}$: If $\mu_{iak} = 1$, Yellow tag defect $k'$ can be rectified by choosing repair activity $a$ for section $i$
- $W^\text{min}_t$: Minimal allowed work time at time period $t$
- $W^\text{max}_t$: Maximal allowed work time at time period $t$
The following model decision variables are defined:

\[ x_{iat} \quad \forall i, a, t \]

binary variables that denote whether repair activity \( a \) is chosen for section \( i \) at time step \( t \) (\( x_{iat} = 1 \)), or not (\( x_{iat} = 0 \))

\[ y_{lt} \quad \forall l, t \]

binary variables that denote whether defect \( l \) is the center of a cluster to be rectified at time step \( t \) (\( y_{lt} = 1 \)), or not (\( y_{lt} = 0 \))

\[ z_{kt} \quad \forall k, t \]

binary variables that denote whether defect \( k \) is rectified at time step \( t \) (\( z_{kt} = 1 \)), or not (\( z_{kt} = 0 \))

\[ w_{kl} \quad \forall k, l \]

binary variables that denote whether defect \( k \) is rectified within a cluster centered at \( l \) at time step \( t \) (\( w_{kl} = 1 \)), or not (\( w_{kl} = 0 \))

\[ B \]

Budget including both rectification costs and travel costs

(1) Objective

The optimization model denoted as the Cost-based Formulation (CF) can now be formulated as follows:

\[ (\text{CF}) \quad \text{Min} \quad B + \sum_{i} \sum_{a} \sum_{t} x_{iat} C_d P_{ia} (\Delta t) \]

Objective (3a) aims to minimize the total expected cost, which is the sum of budget and derailment costs in the inspection interval. Budget is further defined in Constraint (3b), composed of rectification costs and travel costs. The merit of treating budget as a decision variable is that the optimization model can produce the best needed budget to achieve system optimality. If budget is given and limited, the objective can be modified as below,

\[ \text{Min} \quad \sum_{k} \sum_{l} w_{kl} C_{kl} + \sum_{k} \sum_{t} z_{kt} C_{k} + \sum_{i} \sum_{a} \sum_{t} x_{iat} C_d P_{ia} (\Delta t) \]

Note that travel costs and rectification costs are for both Red and Yellow tag defects, while derailment costs are only caused by Yellow tags, since Red tags are always repaired soon after they are discovered. The rectification of Red tags is considered in the formulation since this affects the total travel costs.

(2) Constraints

i. Budget constraints:

\[ \sum_{k} \sum_{l} w_{kl} C_{kl} + \sum_{k} \sum_{t} z_{kt} C_{k} \leq B \]

If budget is given, constraints (3b) bound the allowable total repair costs. Otherwise, the constraints calculate needed budget when the inequality will be tight.

ii. General rectification constraints:

\[ \lambda_{kt} \leq \sum_{t} z_{kt} \leq 1 \quad \forall k \in K_{a} \quad (3c) \]

\[ \sum_{t} t_{kl} \leq L_{k} \quad \forall k \in K_{a} \quad (3d) \]
Constraints (3c) ensure that each defect can only be rectified once, and all the Red tags and a portion of severe Yellow tags ($P^R_k(\Delta t) \geq \delta$) should be chosen in the repair horizon. Constraints (3d) guarantee that the defect will be rectified before its due date. After a regular inspection run, each Red tag will be assigned a repair due date, from a couple of days up to a few weeks, according to its severity. In this paper, the due periods $L_k$ for Yellow tags are obtained as the minimal values to satisfy inequality $P^R_k(L_k) \geq \delta$.

iii. Relation constraints among decision variables:

\[
\begin{align*}
&z_{kl} \leq \sum_t y_{lt} \quad \forall k \in K, t \in \{1, ..., T_R\} \\
y_{lt} \leq z_{lt} \quad \forall l \in K, t \in \{1, ..., T_R\} \\
w_{kl} \leq \sum_t y_{lt} \quad \forall k, l \in K \\
y_{lt} \leq \sum_k w_{kl} \quad \forall l \in K, t \in \{1, ..., T_R\} \\
z_{kl} \leq \sum_t w_{lt} \quad \forall k \in K, t \in \{1, ..., T_R\}
\end{align*}
\]

Constraints (3e) and (3f) describe the relationship between $z_{kl}$ and $y_{lt}$. $z_{kl}$ can be 1 only if any defect is rectified at time $t$, that is: $\sum_t y_{lt} = 1$. And $y_{lt} = 1$ will lead to $z_{kl} = 1$. Similarly, Constraints (3g) and (3h) specify the relationship between $y_{lt}$ and $w_{kl}$. Constraints (3i) ensure that $z_{kl}$ can be 1 only if defect $k$ is rectified within any cluster, that is: $\sum_t w_{lt} = 1$.

iv. Cluster repair constraints:

\[
\begin{align*}
&\sum_l y_{lt} \leq 1 \quad \forall l \in K \\
&\sum_t y_{lt} \leq 1 \quad \forall t \in \{1, ..., T_R\}
\end{align*}
\]

Constraints (3j) and (3k) ensure that a defect can only be the center of at most one cluster and each time step has at most one cluster, respectively.

v. Yellow tag rectification constraints:

\[
\begin{align*}
&\sum_a \sum_t x_{iat} = 1 \quad \forall i \in I \\
&\mu_{iak} x_{iat} \leq z_{kt} \quad \forall k \in K, i \in I, a \in A, t \in \{1, ..., T_R\} \\
z_{kt} \leq \sum_i \sum_a \mu_{iak} x_{iat} \quad \forall k \in K, t \in \{1, ..., T_R\}
\end{align*}
\]
According to constraints (3l), each track section containing Yellow tags should be assigned to an activity at each time step. Activities are defined as a combination of defect types to be repaired at a track section. Grouping the Yellow tags together for rectification seems reasonable for the following two practical reasons: 1) the work team doesn’t need to switch repair equipment and materials while handling multiple defects of the same type, thereby reducing the potential total costs, and 2) it is compatible with the proposed derailment risk model, which is established based on aggregated value for certain defect types rather than individual defects. Suppose that a section is observed to contain Yellow tags of 3 defect categories. In this case, there are $2^3 = 8$ activities available to the decision maker, because the local track master can choose to repair none, single type, mixed types, or all of the defects of each type in any possible combination. Constraints (3m) and (3n) build the relationship between individual defect repair $z_{kt}$ and section repair $x_{iat}$. In (3m), applying a repair activity on a section will rectify all the associated defects. Constraints (3n) guarantee that each chosen Yellow tag corresponds to a repair activity.

**vi. Work time constraints:**

$$W_t \min \sum_i y_{li} \leq \sum_k \tau_k z_{klt} \leq W_t \max \quad \forall l \in K_a, t \in \{1, ..., T_R\}$$

Constraints (3o) implement work time limitations for each rectification period.

The optimization model is a mixed integer program (MIP) which is linear in the objective function as well as constraints, and it can be solved using standard commercial solvers.

Note that the cost of derailment $C_D$ is a random variable. Figure 4 presents the histogram of the derailment cost data and a fitted probability density function. The derailment cost distribution follows a strongly heavy-tailed distribution, ranging from hundreds of US dollars to millions of US dollars. The expected derailment cost based on our data set is around $510K$.

Due to the structure of our proposed formulation, $C_D$ is only present in the objective function. When trying to minimize the total expected costs, we can simply replace derailment cost $C_D$ with the expected derailment cost $\overline{C}_D$ in the model. This is because the total expected cost is given as:

$$E\left[ B + \sum_i \sum_a \sum_t x_{iat} C_{nia} P_{ia}^D(\Delta t) \right] = B + \sum_i \sum_a \sum_t x_{iat} E[C_D] P_{ia}^D(\Delta t) = B + \sum_t \sum_a \sum_i x_{iat} \overline{C}_D P_{ia}^D(\Delta t)$$

(3p)
since the expectation of a sum is the sum of expectations. According to expression (3p), the
stochastic objective will eventually be derived in the same structure as the deterministic one
(3a-1). Therefore, the proposed MIP handles uncertain derailment costs.
Instead of minimizing total expected costs, an alternate objective is to minimize a function
of the maximal derailment probability across all track sections.

$$\text{Min } \{ \beta B + \sum_a \sum_t x_{ait} \Delta P(a, t) \}$$

where $\beta$ denotes a weighting factor, which should be sufficiently small so as to avoid
overwhelming the derailment risk. Such min-max objectives try to hedge the worst-case track
condition in terms of safety and increase the robustness of track infrastructure from a systems
point of view. We denote this model as the Risk-based Formulation (RF), described as
follows:

(RF) $\text{Min } \beta B + F$ \hspace{1cm} (4a)

Subject to

$$\sum_a \sum_t x_{ait} \Delta P(a, t) \leq F \quad \forall i \in I$$ \hspace{1cm} (4b)

And Constraints (3b)-3(o)

where $F$ represents the maximal derailment probability after rectification.

6 Numerical Examples

In this study, the proposed rectification model is applied to geo-defects detected from
inspection runs occurring in December 2011. Among U.S. nation-wide line segments, five
main line freight tracks denoted A, B, C, D, and E, ranging from 50 miles to 500 miles, are
selected as case studies of track rectification planning (see Table 4). In the preprocessing
stage, each track segment is further divided into no more than 2 mile sections according to
network structure and traffic data. Red tags are rare events, while the occurrences of Yellow
tags are much larger. On average, we can detect Yellow tags within every 1~2 miles.
However, the number of Yellow tags is not evenly distributed through the track. After a
complete geometry inspection run, the track sections with defects only account for 15%~20%
of total sections, and among these track sections, each contains around 3~5 defects. This
means that the derailment risk is not uniformly distributed in the entire track. Therefore, for
purposes of efficient Yellow tag repair, it seems reasonable to consider rectifying sections
rather than individual tags. Note that the number of binary variables and number of constraints
increase exponentially when the size of the track increases from around 50 miles to 500 miles.
Track segments longer than 500 miles are therefore not considered in this paper.

The number of Yellow tags are aggregated at the section level in order to build the set of repair activities as well as parameters $\mu_{\text{track}}$. Considering a 90 day inspection interval $\Delta t$, we obtain the derailment probability $P_{\text{D}}(\Delta t)$ before the next inspection run for each section and each repair activity based on the proposed Cox PH model. By applying the defect deterioration model, we calculate the probability $P_{\text{R}}(\Delta t)$ of Yellow tags converting to Red. If $P_{\text{R}}(\Delta t)$ is greater than a predefined threshold $\delta$, the corresponding Yellow defect $k'$ will be labeled as “must repair”. The rectification horizon $T_k$ is set to be less than 22, representing typical repair due in work days of a month. Due to lack of data, the Red tag repair due periods $L_k k \in K_a \setminus K_r$ are randomly sampled from a uniform distribution in the interval $[1, 22]$, and the Yellow repair due periods $L_k k \in K_r$ are obtained as the minimal values that satisfy inequality $P_{\text{R}}(L_k) \geq \delta$.

Accurately estimating the exact repair costs is very difficult, due to variations in equipment, fleet and personnel at each local maintenance center. For simplicity, we only take into account three different defect groups and assume that costs for both Yellow and Red tags stay constant across different maintenance scenarios, as shown in Table 5. In this paper, we first take the average of rectification costs for Red tags, and then scale by 0.9 to obtain the corresponding Yellow tag rectification costs.

All five rectification optimization instances were solved by CPLEX 12.3 on a personal computer with a Quad-Core 3.5GHz CPU and 16 GB RAM. Table 6 summarizes the computational results of rectification planning across five tracks, with both the CF and RF optimization models. As shown in Table 6, we could obtain exact optimal integer solutions for A, B, C, D within a few minutes. CF model for Track E took about 900 seconds to be solved and no exact optimal integer solutions could be found for RF model of track E within 1 hour. However, comparing the best feasible solutions (obtained within 1 hour) with a continuous linear programming objective function value that provides an upper bound for the original discrete optimization problem, the gap is found to be less than 0.01%. This implies that our current best solution may be close enough to the exact optimal integer solution.

Note that the two models share most of the constraints in common, but have different objectives: CF reduces total expected costs, whereas RF minimizes the maximal possible derailment risk. Compared with the CF solutions, RF requires more budget for short and
medium tracks (A, B, and C), aiming to repair track sections with high derailment risk. For longer tracks (D and E), both CF and RF provide similar values for the optimal budget. One reason could be that the probability of observing track sections with “high base risk”, representing the risk after rectify all the defects, is much higher for long tracks. However, RF minimizes the worse situation for track sections with high derailment risks. In that sense, RF is more conservative than CF, but it may be more suitable for users requiring a more risk-averse maintenance strategy.

To further examine the differences between CF and RF solutions, we present the details of track repair activity solutions for track A, shown in Table 7. Each track has $2^n$ repair activities to be conducted, where $n$ is the number of defect types. For each repair activity $a$ at section $i$, derailment probability $P^D_{ia}(\Delta t)$ is obtained using equation (2d). Rectification costs of a repair activity at each section are also provided in Table 7. CF aims to balance the costs from derailment and rectification, while RF considers minimizing the maximal derailment probability as the top priority. For example, section 1235 has the highest derailment probability 0.0145. The repair activity “GAGE_W1” is chosen in RF model since it reduces the probability from 0.0145 to 0.0088. However, no repair activity is applied for section 1235 in the CF model, due to its high rectification cost (over $15,000). To illustrate the deterioration probability from Yellow to Red, we also present the maximal deterioration probability across each track $i$. We set the threshold $\delta$ as 0.85. If the deterioration probability for any Yellow tag in a section is higher than 0.85, tags of the same type are labeled as “must repair” tags, e.g. XLEVEL in section 1227.

For the sake of comparison, we generate severity-based manual plans to mimic the rectification strategy in practice. We define the empirical severity index as:

$$S_{ij} = \exp\left(\frac{\max(\pi_{kj})}{\pi_{j}^{95th}}\right) \ln \left(n_{ij}\right) \quad \forall i, j$$

where $\pi_{kj}$ is the amplitude of Yellow tag defect $k$ in section $i$ and type $j$, $\pi_{j}^{95th}$ denotes the 95th percentile of amplitude of geo-defect type $j$, and $n_{ij}$ represents the number of Yellow tag defects in section $i$ and type $j$. Given the same amount of budget specified by the CF model, severity measures $S_{ij}$ are ranked and selected in descending order, until the budget is consumed. Then the selected sections are simply assigned to their repair due periods, and further adjusted to satisfy the work time limit at each period. Finally, considering additional travel costs, sections with lowest $S_{ij}$ are excluded until the budget constraint is satisfied again.
Figure 5 compares the total costs of CF models to the manual plan for the five track segments. The percentage changes of total costs from the manual plan as compared to the optimal CF model are labeled at the top of the bars in the bar chart. As the rectification problem gets larger in scale from track A to E, the percentage of reduced total costs gained by applying CF against the manual plan increases dramatically, from 12.2% to 35.4%. On average, the CF model reduces total costs by around 20%.

7 Concluding Discussion

This paper presents an analytical framework to address the track geo-defect rectification problem by integrating three models, (i) a statistical deterioration model to predict the amplitude of Yellow tag defects in the future, (ii) a track derailment risk model to dynamically predict the derailment risk as a function of the different types of geo-defects, and (iii) an optimization model to make geo-defect rectification decisions with two different objectives: a cost-based formulation (CF) and a risk-based formulation (RF).

According to our data-driven study, different geo-defect types deteriorate at different speeds. Most of them exhibit larger deterioration rates with higher traffic, but some of them are found not to be very sensitive to traffic data. Track derailment risk is associated with each section of track in 2 miles and each type of geo-defects. We observe that the number of defects in each section is very significant for some types of geo-defects and that the 90 percentiles of geo-defect amplitudes play an important role in predicting derailment risk.

Comparing the CF and RF optimization models, CF aims to reduce total expected costs, while RF minimizes the maximal derailment risk. Risk-averse track local masters may wish to adopt RF model since it hedges against the worst-case situation of derailment risk at each section. The CF model may be more suitable if the railroad company would prefer to maintain a certain level of total costs. Real-world case studies on an existing US railway network show that the proposed methodology yields reliable and economical solutions, as compared to current rail industry practices.

The rectification problem involves optimizing a large number of integer decision variables and constraints, which is difficult to solve in general. We tackle this issue by assuming that Yellow tags are rectified in bulk, at the section and defect type level. Based on the computational results from our numerical case studies, we learn that we can obtain an exact optimal solution within a short span of time, such as a few minutes, in cases where the railway track is up to a couple of hundred miles long. This is therefore able to handle most tracks operated by a local track master. In cases where we cannot find an exact optimal solution
within e.g. 1 hour, the best feasible solution obtained is acceptable since its corresponding
objective function value lies close to the upper bound, as computed by the relaxed continuous
linear programming problem.

Our proposed model assumes that the budget is a decision variable. Instead, we could also
consider the case where the budget is fixed and provided to the user. Track possession costs
are usually one of the biggest concerns while conducting railway preventive maintenance
activities. However, in corrective maintenance activities, safety is always the top priority. In
this paper, we consider track possession costs indirectly by minimizing the travel distance
between defects in each repair period. In future work, one may also wish to consider repair job
scheduling and crew scheduling based on existing train timetables to minimize travel, as well
as handling and possession costs. As a final remark, we would like to note that our track
rectification model could be easily extended to solve other practical problems, particularly
those involving further constraints that a local track master may need to take into account.

References

Ahern, A. & Anandarajah, G., 2007. Railway Projects Prioritisation for Investment:

TRITA-INFRA 02-035. Department of Infrastructure, Royal Institute of Technology.

Track Geometry. Available at: http://www.railroadtraining.biz/a-c/subpart_c.htm


Quality through Different Methodologies. International Journal of Technology, (1),
pp.38–47.


Decision Methodologies Used in Capital Investment Decision Analysis. Engineering
Economist, 42, pp.179–201.

Budai, G., Huisman, D. & Dekker, R., 2006. Scheduling Preventive Railway Maintenance


Figure 1 Some track geometry parameters (ARC-TECH.NET 2012)

Figure 2 Flow chart comparing the existing and proposed decision models for geo-defect rectification.
Figure 3 Cumulative hazard of Cox-Snell residuals

Figure 4 Probability density of derailment cost
Figure 5 Comparisons of the total cost of CF and manual solutions
<table>
<thead>
<tr>
<th>Defect Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALIGN</td>
<td>ALIGN is the projection of the track geometry of each rail or the track center line onto the horizontal plane, also known as “straightness” of the tracks.</td>
</tr>
<tr>
<td>CANT</td>
<td>Rail cant (angle) measures the amount of vertical deviation between two flat rails from their designed value. (1 degree = 1/8” for all rail weights, approximately)</td>
</tr>
<tr>
<td>DIP</td>
<td>DIP is the largest change in elevation of the centerline of the track within a certain moving window distance. Dip may represent either a depression or a hump in the track and approximates the profile of the centerline of the track.</td>
</tr>
<tr>
<td>GAGE_C</td>
<td>Gage Change is the difference in two gage readings in a certain distance measurement interval.</td>
</tr>
<tr>
<td>GAGE_TGHT</td>
<td>GAGE_TGHT measures how much tighter from standard gage (56-1/2”).</td>
</tr>
<tr>
<td>GAGE_W1</td>
<td>Gage is the distance between right and left rail measured 5/8” below the railhead. GAGE_W1 for wood ties (sleepers) measures how much wider from standard gage (56-1/2”). The amplitude of GAGE_W1 plus 56-1/2” is equal to the actual track gage reading.</td>
</tr>
<tr>
<td>GAGE_W2</td>
<td>Same as GAGE_W1, but for concrete ties.</td>
</tr>
<tr>
<td>HARM_X</td>
<td>Harmonic cross-level defect is two cross-level deviations a certain distance apart in a curve.</td>
</tr>
<tr>
<td>OVERELEV</td>
<td>Over-elevation occurs when there is an excessive amount of elevation in a curve (overbalance) based on the degree of curvature and the board track speed.</td>
</tr>
<tr>
<td>REV_X</td>
<td>Reverse cross-level occurs when the right rail is low in a left-hand curve or the left rail is low in a right-hand curve.</td>
</tr>
<tr>
<td>SUPER_X</td>
<td>Super cross-level is cross-level, elevation or super-elevation measured at a single point in a curve.</td>
</tr>
<tr>
<td>SURF</td>
<td>Uniformity of rail surface measured in short distances along the tread of the rails. Rail surface is measured over a 62-foot chord, the same chord length as the FRA specification.</td>
</tr>
<tr>
<td>TWIST</td>
<td>Twist is the difference between two cross-level measurements a certain distance apart.</td>
</tr>
<tr>
<td>WARP</td>
<td>Warp is the difference between two cross-level or elevation measurements up to a certain distance apart.</td>
</tr>
<tr>
<td>WEAR</td>
<td>The Automated Rail Weight Identification System (ARWIS) identifies the rail weight while the car is testing and measures the amount of head loss. The system measures for vertical head wear (VHW) and gage face wear (GFW) per rail.</td>
</tr>
<tr>
<td>XLEVEL</td>
<td>Cross-level is the difference in elevation between the top surfaces of the rails at a single point in a tangent track segment.</td>
</tr>
</tbody>
</table>
Table 2 Parameters for the track deterioration model (for selected defect types)

<table>
<thead>
<tr>
<th>Defect Type</th>
<th>$\alpha_0$ - intercept</th>
<th>$\alpha_1$ - Traffic (MGT)</th>
<th>$\alpha_2$ - Traffic (number of cars)</th>
<th>$\alpha_3$ - Traffic (number of trains)</th>
<th>$\alpha_4$ - Sequence number</th>
<th>Mean squared error (MSE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CANT</td>
<td>-7.66</td>
<td>-7.01E-02</td>
<td>6.05E-06</td>
<td>6.52E-04</td>
<td>0.067</td>
<td>0.242</td>
</tr>
<tr>
<td>DIP</td>
<td>-7.58</td>
<td>7.21E-02</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>0.099</td>
</tr>
<tr>
<td>GAGE_C</td>
<td>-8.53</td>
<td>8.62E-02</td>
<td>--</td>
<td>--</td>
<td>0.113</td>
<td>0.046</td>
</tr>
<tr>
<td>GAGE_W1</td>
<td>-7.19</td>
<td>3.55E-02</td>
<td>4.59E-06</td>
<td>--</td>
<td>0.068</td>
<td>0.021</td>
</tr>
<tr>
<td>GAGE_W2</td>
<td>-8.08</td>
<td>1.90E-02</td>
<td>2.05E-04</td>
<td>0.080</td>
<td>0.061</td>
<td></td>
</tr>
<tr>
<td>OVERELEV</td>
<td>-7.58</td>
<td>2.45E-01</td>
<td>--</td>
<td>2.05E-04</td>
<td>0.080</td>
<td></td>
</tr>
<tr>
<td>SURF</td>
<td>-6.99</td>
<td>2.00E-01</td>
<td>--</td>
<td>-1.33E-03</td>
<td>0.044</td>
<td>0.007</td>
</tr>
<tr>
<td>WEAR</td>
<td>-8.22</td>
<td>2.95E-02</td>
<td>--</td>
<td>4.73E-04</td>
<td>0.075</td>
<td>0.080</td>
</tr>
<tr>
<td>XLEVEL</td>
<td>-7.66</td>
<td>--</td>
<td>2.64E-06</td>
<td>3.23E-04</td>
<td>0.092</td>
<td>0.019</td>
</tr>
</tbody>
</table>

*: “--” means that the coefficient is not significant.
| covariates     | coef      | exp(coef) (hazard ratio) | se(coef) | z     | Pr(>|z|) |
|---------------|-----------|-------------------------|----------|-------|----------|
| numYEL_GAGE_W1| 1.01E-01  | 1.11E+00                | 1.96E-02 | 5.165 | 2.40E-07 |
| numYEL_GAGE_W2| 2.28E-01  | 1.26E+00                | 1.17E-01 | 1.948 | 0.05139  |
| numYEL_REV_X  | 6.66E-01  | 1.95E+00                | 3.21E-01 | 2.077 | 0.03782  |
| amp90_ALIGN   | 4.17E+00  | 6.45E+01                | 2.40E+00 | 1.736 | 0.08261  |
| amp90_DIP     | 9.58E-01  | 2.61E+00                | 4.67E-01 | 2.052 | 0.04016  |
| amp90_GAGE_C  | 6.33E-01  | 1.88E+00                | 2.10E-01 | 3.013 | 0.00259  |
| amp90_HARM_X  | 5.43E+00  | 2.29E+02                | 2.78E+00 | 1.958 | 0.05028  |
| amp90_SUPER_X | 5.10E-01  | 1.67E+00                | 1.68E-01 | 3.028 | 0.00246  |
| amp90_WARP    | 7.44E-01  | 2.10E+00                | 3.69E-01 | 2.014 | 0.04404  |
| amp90_WEAR    | 1.54E+00  | 4.65E+00                | 8.34E-01 | 1.843 | 0.06533  |
Table 4 Summary of test track segments

<table>
<thead>
<tr>
<th>Track name</th>
<th>Length (mile)</th>
<th>Num. of sections with defects</th>
<th>Num. of defect types</th>
<th>Num. of Yellow tags</th>
<th>Num. of Red tags</th>
<th>Rectification horizon $T^r_k$</th>
<th>Num. of binary variables</th>
<th>Num. of constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>50</td>
<td>13</td>
<td>3</td>
<td>60</td>
<td>4</td>
<td>5</td>
<td>4,886</td>
<td>14,974</td>
</tr>
<tr>
<td>B</td>
<td>100</td>
<td>21</td>
<td>6</td>
<td>77</td>
<td>2</td>
<td>10</td>
<td>8,441</td>
<td>58,300</td>
</tr>
<tr>
<td>C</td>
<td>200</td>
<td>34</td>
<td>7</td>
<td>100</td>
<td>18</td>
<td>15</td>
<td>18,760</td>
<td>149,090</td>
</tr>
<tr>
<td>D</td>
<td>300</td>
<td>49</td>
<td>8</td>
<td>122</td>
<td>37</td>
<td>20</td>
<td>34,801</td>
<td>426,754</td>
</tr>
<tr>
<td>E</td>
<td>500</td>
<td>94</td>
<td>8</td>
<td>343</td>
<td>38</td>
<td>22</td>
<td>157,185</td>
<td>1,157,160</td>
</tr>
</tbody>
</table>
Table 5 Average rectification cost for individual geo-defects

<table>
<thead>
<tr>
<th>Geo-defect types</th>
<th>Average rectification cost for a Yellow tag defect (USD)</th>
<th>Average rectification cost for a Red tag defect (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gage related: GAGE_W1, GAGE_W2 and GAGE_C</td>
<td>1539</td>
<td>1710</td>
</tr>
<tr>
<td>Surface related: DIP and SURF</td>
<td>1125</td>
<td>1250</td>
</tr>
<tr>
<td>Other defect types</td>
<td>1534.5</td>
<td>1705</td>
</tr>
</tbody>
</table>
Table 6 Summary of computational results

<table>
<thead>
<tr>
<th>Track name</th>
<th>CF (USD)</th>
<th>RF (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>19,136.29</td>
<td>37,612.71</td>
</tr>
<tr>
<td>B</td>
<td>23,475.49</td>
<td>60,373.72</td>
</tr>
<tr>
<td>C</td>
<td>94,947.81</td>
<td>112,029.82</td>
</tr>
<tr>
<td>D</td>
<td>117,148.79</td>
<td>111,465.19</td>
</tr>
<tr>
<td>E</td>
<td>212,250.15</td>
<td>208,995.63</td>
</tr>
</tbody>
</table>

Optimal budget (USD)

<table>
<thead>
<tr>
<th>Track name</th>
<th>Optimal objective</th>
<th>Maximal derailment probability</th>
<th>Solution time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>64,141.26</td>
<td>0.0125</td>
<td>2.15</td>
</tr>
<tr>
<td>B</td>
<td>127,086.22</td>
<td>0.01976</td>
<td>4.46</td>
</tr>
<tr>
<td>C</td>
<td>186,058.50</td>
<td>0.0121</td>
<td>14.087</td>
</tr>
<tr>
<td>D</td>
<td>291,994.38</td>
<td>0.0279</td>
<td>54.27</td>
</tr>
<tr>
<td>E</td>
<td>493,349.31</td>
<td>0.0268</td>
<td>895.25</td>
</tr>
</tbody>
</table>

*: CPLEX didn’t give optimal solutions in 3600 seconds.
Table 7 Solutions to rectification optimization for Track A

<table>
<thead>
<tr>
<th>Section id ($i$)</th>
<th>Repair activity ($a$)</th>
<th>$P_a^D(\Delta t)$</th>
<th>Rectification costs of a activity</th>
<th>Maximal probability of Yellow tags deteriorating to Red</th>
<th>$\sum r x_{iat}$</th>
<th>$\sum r x_{iat}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1216</td>
<td>NONE</td>
<td>0.0099</td>
<td>0</td>
<td>0.457</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1216</td>
<td>GAGE_W1</td>
<td>0.0058</td>
<td>4,617</td>
<td>0.457</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1217</td>
<td>NONE</td>
<td>0.0050</td>
<td>0</td>
<td>0.254</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1217</td>
<td>GAGE_W1</td>
<td>0.0035</td>
<td>10,773</td>
<td>0.254</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1218</td>
<td>NONE</td>
<td>0.0047</td>
<td>0</td>
<td>0.076</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1218</td>
<td>GAGE_W1</td>
<td>0.0045</td>
<td>1,539</td>
<td>0.076</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1225</td>
<td>NONE</td>
<td>0.0057</td>
<td>0</td>
<td>0.277</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1225</td>
<td>XLEVEL</td>
<td>0.0057</td>
<td>3,069</td>
<td>0.277</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1227</td>
<td>NONE</td>
<td>0.0072</td>
<td>0</td>
<td>0.974</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1227</td>
<td>XLEVEL</td>
<td>0.0072</td>
<td>3,069</td>
<td>0.974</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1232</td>
<td>NONE</td>
<td>0.0054</td>
<td>0</td>
<td>0.991</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1232</td>
<td>GAGE_W1</td>
<td>0.0044</td>
<td>6,156</td>
<td>0.991</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1234</td>
<td>NONE</td>
<td>0.0063</td>
<td>0</td>
<td>0.648</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1234</td>
<td>GAGE_W1</td>
<td>0.0045</td>
<td>10,773</td>
<td>0.648</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1235</td>
<td>NONE</td>
<td>0.0145</td>
<td>0</td>
<td>0.268</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1235</td>
<td>GAGE_W1</td>
<td>0.0088</td>
<td>15,390</td>
<td>0.268</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1235</td>
<td>XLEVEL</td>
<td>0.0145</td>
<td>4,603.5</td>
<td>0.268</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1235</td>
<td>GAGE_W1 + XLEVEL</td>
<td>0.0088</td>
<td>19,993.5</td>
<td>0.268</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1237</td>
<td>NONE</td>
<td>0.0064</td>
<td>0</td>
<td>0.378</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1237</td>
<td>GAGE_W1</td>
<td>0.0043</td>
<td>12,312</td>
<td>0.378</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1238</td>
<td>NONE</td>
<td>0.0076</td>
<td>0</td>
<td>0.446</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1238</td>
<td>GAGE_W1</td>
<td>0.0045</td>
<td>13,851</td>
<td>0.446</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1240</td>
<td>NONE</td>
<td>0.0109</td>
<td>0</td>
<td>0.382</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1240</td>
<td>GAGE_C</td>
<td>0.0061</td>
<td>1,539</td>
<td>0.382</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1240</td>
<td>GAGE_W1</td>
<td>0.0091</td>
<td>1,539</td>
<td>0.382</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1240</td>
<td>GAGE_C + GAGE_W1</td>
<td>0.0051</td>
<td>3,078</td>
<td>0.382</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1241</td>
<td>NONE</td>
<td>0.0084</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1241</td>
<td>GAGE_C</td>
<td>0.0048</td>
<td>1,539</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1242</td>
<td>NONE</td>
<td>0.0046</td>
<td>0</td>
<td>0.049</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1242</td>
<td>GAGE_W1</td>
<td>0.0044</td>
<td>1,539</td>
<td>0.049</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>